AD NUMBER

AD476507

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; Oct 1965. Other requests shall be referred to Army Research Office, Research Triangle Park, NC.

AUTHORITY

USARO ltr, 21 Nov 1966

ARO-D Report 65-3

PROCEEDINGS OF THE TENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by

The Army Mathematics Steering Committee

on Behalf of

THE OFFICE OF THE CHIEF OF RESEARCH AND DEVELOPMENT

U. S. ARMY RESEARCH OFFICE-DURHAM

Report No. 65-3
October 1965

PROCEEDINGS OF THE TENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH DEVELOPMENT AND TESTING

Sponsored by the Army Mathematics Steering Committee

Host

The Army Research Office, Office Chief of Research and Development
Department of the Army
Washington, D. C.
4-6 November 1964

U. S. Army Research Office-Durham
Box CM, Duke Station
Durham, North Carolina

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.
- Pages with background color or light colored printing.
- Pages with small type or poor printing; and or
- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☑ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd) Page

* This paper was presented at the conference. It does not appear in these Proceedings.

TABLE OF CONTENTS (cont'd) Page

---

* This paper was presented at the conference. It does not appear in these Proceedings.

TABLE OF CONTENTS (cont'd) Page

* This paper was presented at the conference. It does not appear in these Proceedings.

TABLE OF CONTENTS (cont'd)

* We anticipate that this paper will appear in the Proceedings of the Eleventh Conference on the Design of Experiments

FOREWORD

The Army Research Office, Office of the Chief of Research and Development, Department of the Army, served as host for the Tenth Conference on Design of Experiments in Army Research, Development and Testing. The Conference was held in Washington, D. C. during 4-6 November 1964.

The continued success of these conferences is a tribute to the foresight of Professor Samuel S. Wilks who conceived the idea of holding such conferences and chaired the Program Committee for the first nine conferences. Unfortunately, due to his untimely death, Professor Wilks could not participate in this Tenth Anniversary Conference. His effort in connection with these Conferences was only one of Professor Wilks' many contributions to the Army. His wise counsel and advice will be missed. As a small recognition for his services to the Army, this Tenth Anniversary Conference was dedicated to the memory of Professor Wilks.

Almost 300 statisticians, engineers and physicists from the Army, other government agencies, Army contractors, and universities attended the conference. This number far exceeds the attendance at any of the previous conferences and reflects, in part, the esteem for Professor Wilks in the statistical community.

One surprising feature was the announcement that Mr. Philip G. Rust of Thomasville, Georgia, had contributed funds for a Samuel S. Wilks Award to be presented annually at the Design of Experiments Conference. It is especially gratifying that a long-time civilian employee of the U. S. Army, Dr. Frank Grubbs, Associate Technical Director of the Ballistic Research Laboratories, was the recipient of the initial award. We are appreciative that the American Statistical Association has accepted the responsibility for determining future Award winners.

Because of the particular significance of this Tenth Conference, the Program Committee invited several distinguished statisticians to deliver papers: Professor H. O. Hartley, Professor Oscar Kempthorne, Dr. M G. Kendall and Professor John W. Tukey. Professor Gerald J. Lieberman served as chairman of the Panel Discussion on Regression Analysis and arranged for Professor G.E. P. Box, Professor Jack C. Kiefer, and Professor Ingram Olkin to give pertinent papers and for

Professor Robert Bechhofer to serve as the invited discussant. In addition to these invited addresses, 15 papers were given in the Clinical Sessions and 18 papers in the Technical Sessions. Additional highlights of the meetings were the after dinner presentations by Dr. Churchill Eisenhart and Dr. W. J. Youden.

It is fitting to give recognition for the particular activities of two groups with regard to these Conferences. The Army Mathematics Steering Committee (AMSC), currently chaired by Dr. I. R. Hershner, Jr. is commended for its strong support of these Conferences because of the actual and potential gains obtained by Army facilities. The members of the Tenth Conference Program Committee are commended for their work in obtaining speakers, selecting a location and planning the overall program. The members of this Committee were: Dr. F. G. Dressel (Secretary), Mr. Fred Frishman, Dr. Walter D. Foster, Dr. Frank E. Grubbs (Chairman), Professor Boyd Harshbarger, Professor H. L. Lucas, Dr. Clifford J. Maloney, Professor Henry B. Mann and Professor Geoffrey S. Watson. Special credit is given to Dr. F. G. Dressel for performing all of the necessary details regarding the program, invitations and the publication of these Proceedings.

It is planned to have an Eleventh Conference at Picatinny Arsenal in 1965. As is well known, these Conferences have been held to assist Army statisticians and their parent organizations. It is hoped that Army statisticians will continue to support these conferences both by the presentation of scientific papers and by their attendance.

WALTER E. LOTZ, JR.
Director of Army Research

TENTH CONFERENCE ON THE DESIGN OF EXPERIMENTS
IN ARMY RESEARCH, DEVELOPMENT AND TESTING

4-6 November 1964

Wednesday, 4 November

0800-0900 REGISTRATION -- Mezzanine Floor in Foyer No. 3 of the Statler-Hilton Hotel

0900-0920 CALLING OF CONFERENCE TO ORDER -- South American Room, Fred Frishman, Chairman on Local Arrangements

0920-1200 GENERAL SESSION 1

Chairman: Major General Austin W. Betts, Deputy Chief of Research and Development

THE STIMULUS OF S. S. WILKS TO ARMY STATISTICS
Major General Leslie E. Simon (Ret'd), Winter Park, Florida

THE SAMUEL S. WILKS AWARD

Announcement: Don Riley, American Statistical Association

Presentation: Philip G. Rust, Thomasville, Georgia

BREAK

DEVELOPMENT OF THE DESIGN OF EXPERIMENTS OVER THE PAST TEN YEARS
Professor Oscar Kempthorne, Iowa State University, Ames, Iowa

1200-1320 LUNCH

Technical Sessions I and II and Clinical Session A will start at 1320 and run to 1500. After a break Technical Sessions III and IV and Clinical Session B will convene at 1540 and run to 1710.

1320 - 1500 TECHNICAL SESSION I -- New York Room

Chairman: W. H. Ewart, Research and Development Division, Army Missile Command, Redstone Arsenal, Alabama

APPLICATION OF DIMENSION THEORY TO MULTIPLE REGRESSION ANALYSIS
David R. Howes, U. S. Army Strategy and Tactics Analysis Group, Bethesda, Maryland

THE USE OF REGRESSION ANALYSIS FOR CORRECTING OF MATRIX EFFECTS IN THE X-RAY FLUORESCENCE ANALYSES OF PYROTECHNIC COMPOSITIONS
R. H. Myers and B. J. Alley, Virginia Polytechnic Institute, Blacksburg, Virginia, Rep. Redstone Arsenal

1320 - 1500 TECHNICAL SESSION II - South American Room

Chairman: Henry Ellner, Directorate for Quality Assurance, Edgewood Arsenal, Maryland

SAMPLING FOR DESTRUCTIVE OR EXPENSIVE TESTING
Joseph Mandelson, Directorate of Quality Assurance, U. S. Army Edgewood Arsenal, Edgewood Arsenal, Md.

TOTAL SAMPLE STATISTICS FROM SUBSAMPLE STATISTICS
Paul C. Cox, Reliability and Statistics Office, Army Missile Test and Evaluation Directorate, White Sands Missile Range, New Mexico

1320 - 1500 CLINICAL SESSION A -- California Room

Chairman: Ira A. DeArmon, Jr., Operations Research Group, Army Chemical Corps, Edgewood Arsenal, Md.

Panelists:

Dr. Frank E. Grubbs, Army Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

Professor H. C. Hartley, Institute of Statistics, Agricultural and Mechancial College, College Station, Texas

Panelists (cont'd):

Dr. Emil H. Jebe, Institute of Science and Technology, The University of Michigan, Ann Arbor, Michigan

Professor Gerald J. Lieberman, Stanford University, Stanford, California

Professor H. L. Lucas, Institute of Statistics, North Carolina State of the U.N.C., Raleigh, North Carolina

SYSTEM CONFIGURATION PROBLEMS AND ERROR SEPARATION PROBLEMS
Fred S. Hanson, Plan and Operations Directorate, White Sands Missile Range, New Mexico

AN EXPERIMENT IN MAKING TECHNICAL DECISIONS USING OPERATIONS RESEARCH AND STATISTICAL METHODS
Andrew H. Jenkins, U. S. Army Missile Command, Huntsville, Alabama, and Edwin M. Bartee, School of Engineering, University of Alabama

1500 - 1540 BREAK

1540 - 1710 TECHNICAL SESSION III -- New York Room

Chairman: Morris A. Rhian, Operations Research Group, Army Chemical Corps, Edgewood Arsenal, Md.

IMPROVEMENT CURVES: PRINCIPLES AND PRACTICES
Jerome H. N. Selman, Stevens Institute of Technology, Rep. the U. S. Army Munitions Command, Dover, N. J.

THE EFFECT OF VALIDITY, LENGTH, AND SCORE CONVERSION ON A MEASURE OF PERSONNEL ALLOCATION EFFICIENCY
Richard C. Sorenson and Cecil D. Johnson, U. S. Army Personnel Research Office, Washington, D. C.

1540 - 1710 TECHNICAL SESSION IV -- South American Room

Chairman: Joseph R. Lane, Technical Evaluation Office, Army Research Office-Durham, Durham, N. C.

A QUANTITATIVE ASSAY FOR CRUDE ANTHRAX TOXINS
Bertram W. Haines, U. S. Army Biological Labs., Fort Detrick, Frederick, Maryland

AN INVESTIGATION OF THE DISTRIBUTION OF DIRECT HITS ON PERSONNEL BY SELF-DISPERSING BOMBLETS
David M. Moss and Theodore W. Horner, Booz-Allen Applied Research, Inc., Bethesda 14, Maryland
Rep. Biomathematics Division of Fort Detrick, Maryland

1540 - 1710 CLINICAL SESSION B -- California Room

Chairman: Henry A. Dihm, Advanced Systems Laboratory, Army Missile Command, Redstone Arsenal, Alabama

Panelists:

Dr. O. P. Bruno, Surveillance Group, Army Ballistics Research Laboratories, Aberdeen Proving Ground, Md.

Dr. Donald S. Burdick, Duke University, Durham, N. C.

Professor Clyde Y. Kramer, Virginia Polytechnic Institute, Blacksburg, Virginia

Dr. R. L. Stearman, C-E-I-R, Inc., Los Angeles, Calif.

Dr. William Wolman, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland

EXPLOSIVE SAFETY AND RELIABILITY ESTIMATES FROM A LIMITED SIZE SAMPLE
J. N. Ayres, L. D. Hampton and I. Kabik, U. S. Naval Ordnance Laboratory, White Oak, Silver Spring, Maryland

COMPARING THE VARIABILITIES OF TWO TEST METHODS USING DATA FOR SEVERAL POPULATIONS
Manfred W. Krimmer, U. S. Army Ammunition Procurement and Supply Agency, Joliet, Illinois

Thursday, 5 November

Technical Session V and Clinical Session C and D will run from 0830-1010. After the break General Session 2 will convene at 1050. After lunch Technical Sessions VI and VII and Clinical Session E will start at 1300 and end at 1420. The Panel Discussion is scheduled to be conducted from 1450 to 1710. Following the banquet, which starts at 1900, there will be two short talks.

0830-1010 TECHNICAL SESSION V -- South American Room

Chairman: R. H. Myers, Statistical Laboratory, Virginia Polytechnic Institute, Blacksburg, Virginia

CYCLIC DESIGNS
H. A. David and F. W. Wolock, University of North Carolina and Virginia Polytechnic Institute, Rep. Army Research Office-Durham

SOME RESULTS ON THE FOUNDATIONS OF STATISTICAL DECISION THEORY
Bernard Harris, J. D. Church, F. V. Atkinson, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin

0830-1010 CLINICAL SESSION C -- California Room

Chairman: Dr. Erwin L. LeClerg, Biometrical Services Division, U. S. Department of Agriculture, Plant Industry, Beltsville, Maryland

Panelists:

Dr. Walter D. Foster, Biometrics Division, Army Biological Warfare Laboratories, Fort Detrick, Md.

Dr. Samuel W. Greenhouse, Biometrics Branch, National Institute of Mental Health, Bethesda, Maryland

Panelists (cont'd):

Professor Clyde Y. Kramer, Virginia Polytechnic Institute, Blacksburg, Virginia

Professor H. L. Lucas, North Carolina State of the UNC, Raleigh, North Carolina

Dr. Clifford J. Maloney, Division of Biologics Standards, National Institutes of Health, Bethesda, Maryland

DISINFECTION OF AEROSOLIZED PATHOGENIC FUNGI ON LABORATORY SURFACES
Richard H. Kruse, Theron D. Green, Richard C. Chambers and Marian W. Jones, U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland

THE EFFECT OF SNAKE VENOM AND ENDOTOXIN ON CORTICAL ELECTRICAL ACTIVITY
James A. Vick, Henry P. Ciuchta, Edward H. Polley, and James Manthei, Directorate of Medical Research, Chemical Research and Development Laboratories, Edgewood Arsenal, Maryland

COMPUTER ANALYSIS OF RHESUS MONKEY IN VISUAL DISCRIMINATION TESTING
John C. Atkinson, Directorate of Medical Research, Chemical Research and Development Laboratories, Edgewood Arsenal, Maryland

0830-1010 CLINICAL SESSION D -- New York Room

Chairman: Lee W. Green, Jr., Florida Research and Development Center, Pratt and Whitney Aircraft, West Palm Beach, Florida

Panelists:

Professor R. E. Bechhofer, Cornell University, Ithaca, New York

Professor G. E. P. Box, the University of Wisconsin, Madison, Wisconsin

Panelists (cont'd):

Dr. T. W. Horner, Booz-Allen Applied Research, Inc., Bethesda, Maryland

Professor G. J. Lieberman, Stanford University, Stanford, California

Dr. H. B. Mann, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wisconsin

FATIGUE - LIMIT ANALYSES AND DESIGN OF FATIGUE EXPERIMENTS
A. H. Soni and R. E. Little, Oklahoma State University, Stillwater, Oklahoma. Representing Army Research Office-Durham

GETTING REGRESSION ANALYSIS IMPLEMENTED
W. H. Ammann, U. S. Army Aviation Materiel Command, St. Louis, Missouri

1010 - 1050 BREAK

1050 - 1150 GENERAL SESSION 2 -- South American Room

Chairman: Dr. Walter D. Foster, Biometric Div., Army Biological Warfare Labs., Fort Detrick, Frederick, Md.

ASSESSMENT AND CORRECTION OF DEFICIENCES IN PERT
Drs. H. O. Hartley and A. W. Wortham, Institute of Statistics, Texas A and M University, College Station, Texas

1150 - 1300 LUNCH

1300 - 1420 TECHNICAL SESSION VI -- South American Room

Chairman: Leonard Pepper, Concrete Division, U. S. Army Engineer Waterways Experiments Station, Vicksburg, Miss.

TEQUILAP: TEN QUANTITATIVE ILLUSIONS OF ADMINISTRATIVE PRACTICE
Clifford J. Maloney

COMBAT VEHICLE FLEET MANAGEMENT
C. J. Christiansen and Mr. G. F. Cooper, Research Analysis Corporation, McLean, Virginia

1300 - 1420 TECHNICAL SESSION VII -- New York Room

Chairman: Eugene F. Smith, Concrete Division, U. S. Army Waterways Experiment Station, Vicksburg, Miss.

APPLICATION OF STATISTICS TO EVALUATE SWIVEL HOOK TYPE CROSS CHAIN FASTENERS FOR MILITARY APPLICATIONS OF TIRE CHAINS
Otto H. Pfeiffer, Components Research and Development Labs., Army Tank-Automotive Center, Warren, Michigan

SOME FACTORS AFFECTING THE PRECISION OF CO-ORDINATE MEASUREMENTS ON PHOTOGENIC PLATES
Desmond O'Connor, Research and Analysis Division, U. S. Army Engineer Geodesy, Intelligence and Mapping Research and Development Agency, Fort Belvoir, Virginia

1300 - 1420 CLINICAL SESSION E -- California Room

Chairman: Joseph Mandelson, Directorate of Quality Assurance, Edgewood Arsenal, Maryland

Panelists:

Professor Donald S. Burdick, Duke University, Durham, North Carolina

Dr. Bernard Harris, Mathematics Research Center, U. S. Army, University of Wisconsin, Madison, Wis.

Professor Ingram Olkin, Stanford University, Stanford, California

Dr. H. M. Rosenblatt, Statistical Research Division, Bureau of the Census, Washington, D. C.

Professor G. S. Watson, The Johns Hopkins University, Baltimore, Maryland

ERROR ANALYSIS PROBLEMS IN THE ESTIMATION OF SPECTRA
Virginia Tipton, Plans and Operations Directorate, White Sands Missile Range, New Mexico

VALIDATION PROBLEMS OF AN INTERFERENCE PREDICTION MODEL
William B. McIntosh, Army Electronics Proving Ground, Fort Huachuca, Arizona

1420 - 1450 BREAK

1450 - 1710 GENERAL SESSION 3 -- South American Room

PANEL DISCUSSION ON REGRESSION ANALYSIS
Chairman: Professor Gerald J. Lieberman, Stanford University

Panelists and the Titles of their Addresses:

USE AND ABUSE OF REGRESSION
Professor G. E. P. Box, The University of Wisconsin

OPTIMUM EXTRAPOLATION AND INTERPOLATION DESIGNS
Professor Jack C. Kiefer, Cornell University

ESTIMATION FOR A REGRESSION MODEL WITH CONVARIANCE
Professor Ingram Olkin, Stanford University

Discussant: Professor Robert Bechhofer, Cornell University

1900 BANQUET

Evening Session Chairman: Dr. I. R. Hershner, Jr., ARO

SAM WILKS AS I REMEMBER HIM
Dr. Churchill Eisenhart, National Bureau of Standards, Washington, D. C.

AN OPERATIONS RESEARCH YARN AND OTHER COMMENTS
Dr. W. J. Youden, National Bureau of Standards, Washington, D. C.

Friday, 6 November

Technical Sessions VIII and IX as well as Clinical Session F run from 0830 to 0950. General Session 4 will start at 1020 and end at 1220.

0830-0950 TECHNICAL SESSION VIII -- South American Room

Chairman: Donald S. Burdick, Duke University, Durham, N. C.

THE DESIGN OF COMPLEX SENSITIVITY EXPERIMENTS
D. Rothman and J. M. Zimmerman, Mathematic and Statistics Group, Rocketdyne, A Division of N. American Aviation, Canoga Park, Calif. Rep. George C. Marshall Space Flight Center, NASA, Huntsville, Alabama

FACTORS AFFECTING SENSITIVITY EXPERIMENTS
J. R. Kniss and W. Wenger, U. S. Army Ballistic Research Labs., Aberdeen Proving Ground, Maryland

0830-0950 TECHNICAL SESSION IX -- New York Room

Chairman: Ralph E. Brown, U. S. Army Munitions Command, Philadelphia, Pennsylvania

A COMPARISON OF RECONNAISSANCE TECHNIQUES FOR LIGHT OBSERVATION HELICOPTERS AND A GROUND SCOUT PLATOON
Harrison N. Hoppes, Barry M. Kibel, Arthur R. Woods, Research Analysis Corporation, McLean, Virginia

A STUDY OF PROBABILITY ASPECTS OF A SIMULATANEOUS SHOCK WAVE PROBLEM
Edward C. Hecht, Nuclear Engineering Directorate, Picatinny Arsenal, Dover, New Jersey

0830-0950 CLINICAL SESSION F -- California Room

Chairman: Dr. B. W. Haines, U. S. Army Biological Laboratories, Fort Detrick, Maryland

Panelists:

Professor R. E. Bechhofer, Cornell University, Ithaca, New York

Mr. David R. Howes, U. S. Army Strategy and Tactics Analysis Group, Bethesda, Maryland

Dr. R. J. Lundegard, Logistics and Mathematical Statistics Branch, Office of Naval Research, Washington, D. C.

Professor Ingram Olkin, Stanford University, Stanford, California

Professor G. S. Watson, The Johns Hopkins University, Baltimore, Maryland

A DATA COLLECTION PROCEDURE FOR ASSESSING NEURO-MOTOR PERFORMANCE IN THE PRESENCE OF MISSILE WOUNDS
William H. Kirby, Jr., William Kokinakis, Larry M. Sturdivan and William P. Johnson, Ballistic Research Aberdeen Proving Ground, Maryland

PROBLEMS IN THE DESIGN OF STATISTICS-GENERATING WAR GAMES
William H. Sutherland, Research Analysis Corporation, McLean, Virginia

0950 - 1020 BREAK

1020 - 1220 GENERAL SESSION 4 -- South American Room

Chairman: Dr. Frank E. Grubbs, Chairman of the Conference, Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland

THE FUTURE OF PROCESSES OF DATA ANALYSIS
Professor John W. Tukey, Princeton University, Princeton, New Jersey

STATISTICS AND MANAGEMENT
Dr. M. G. Kendall, C-E-I-R, London, England

THE STIMULUS OF S. S. WILKS TO ARMY STATISTICS

Leslie E. Simon
Major General, USA (Ret.)

ABSTRACT. The stimulus of S. S. Wilks to the scientific community is discussed briefly, followed by a more detailed account of his originating the idea of a series of Army-wide conferences on design of experiments in Army research, development and testing. The Army's rather satisfactory progress in statistical methodology prior to the conference series is discussed, with comments on its limitations and less than ideal direction of procedure. Wilks' apparent perception of the situation, his courage in undertaking a large and difficult task, and his surprisingly large measure of success is discussed. The importance of carrying on the spirit of Wilks is emphasized, and the creation of The Wilks Award, as a measure to that end is mentioned.

ORIGIN OF THE CONFERENCE SERIES. Mr. Chairman, Fellow Conferees, Ladies and Gentlemen, Samuel Stanley Wilks was my very good friend most of his professional life. Whereas I am aware of many of Wilks' dedicated and outstanding services at a national, if not a world level, I prefer to concentrate my remarks on an area of Wilks' career that is close to home to me: the very valuable services that he did voluntarily for the Army. I am sure that others more able than I will cover his broader services as a teacher, both academic and extra curricular; as a research worker, as an organizer, and as a competent and inspiring leader. Frederick Mosteller has presented an excellent outline of Wilks' worldwide work in the April, 1964 issue of "The American Statistician", under the title, "Samuel S. Wilks; Statesman of Statistics". Mosteller's paper should serve as a guide for other papers on Wilks. However, I cannot help observing that although Mosteller's title is justified, I hope that he will forgive me if I observe that Wilks was by his own choice somewhat lacking in the formality associated with statesmanship. Contrary to one's concept of dignity, Sam was "just folks", whether he was talking with a first-rate scientist, a neophyte in Applied Statistics or a man primarily a soldier. He knew and understood people; and, by nature was ever-ready to give any help within his competence to anyone who genuinely needed it. It was in the latter two capacities, that I had my entree to Wilks.

It was over fifteen years after our initial meeting that Wilks made a proposal that has helped much in improving Army organization, doctrine,

tactics and weapons; and, at the same time contributed to improving the morale of Army personnel, and to saving time and expense in military research and development.

In late 1954 or early 1955, when I was Assistant Chief of Ordnance for Research and Development, U. S. Army, Wilks proposed that the Army establish a series of Army-wide conferences on design of experiments in Army research, development and testing. Dr. Frank E. Grubbs, who, under the authority of my office, had chaired an Ordnance symposium on Statistical Methods in 1953 [1], strongly indorsed Wilks' proposal for Army-wide conferences, devoted primarily to design of experiments; and, of course, I concurred. The Army Mathematics Advisory Panel* (later, designated as the Army Mathematics Steering Committee) operated under the Office of Ordnance Research (now Army Research Office-Durham); and consequently the responsibility for the conferences was assigned to that office. Wilks' proposal was made pursuant to a survey made by the Army Mathematics Steering Committee in which they investigated over 30 Army facilities. They found that one of the most frequently mentioned needs expressed by the scientific personnel was for greater knowledge of modern statistical theory of the design and analysis of experiments. The First Conference on Design of Experiments, in Army Research, Development and Testing was held on October 19-21, 1955 at the Diamond Ordnance Fuze Laboratories and The National Bureau of Standards. Wilks chaired all the conferences up to the present Tenth Conference.

I believe that observing as best we can the time-rate-of-change of the character of these conferences and the concurrent increase of basic understanding of the interrelationships of men, weapons, organization, doctrine, tactics, and research and development, will throw light on the beneficial influence of Wilks on National Defense. I do not mean to infer that all Statistical progress is due to Wilks; but I am sure that much of the progress is due to the spirit of cooperation that he infused, to his influence and to his

---

*The Army Mathematics Advisory Panel, of which Wilks was a member was operated by the Ordnance Corps for the Office of the Chief of Research and Development, U. S. Army. I am indebted to Colonel P. N. Gillon (Ret.), who was both the Commanding Officer of the Office of Ordnance Research (Durham) and the very able Chairman of the Army Mathematics Advisory Panel for the clear, curt minutes and records that he left, and especially for reference [2].

personal contributions. Similarly, I believe that the history of Wilks in this relatively small sub-field of his very active life is a close parallel to the fruitfulness of his activity in other fields to which he devoted far more time. Let us, then, observe the status of Army statistics up to 1953; trace, at least approximately, the conferences on Design of Experiments in Army Research, Development and Testing; and observe the present-day status of Army statistics.

Incidentally, the Army was neither without statistical sophistication in 1953, nor is its knowledge optimum today.

SUMMARY OF ARMY STATISTICAL PROGRESS, BETWEEN WORLD WAR I AND II. Historically, the application of probability theory to the dispersion of shots on a target appears to be about the only Army use of Statistics, prior to World War I. There was a jump in mathematical sophistication during World War I, due to A. A. Bennett [3], Fowler [4], Moulton [5], and others in connection with progress in applying statistics to Ballistic problems. Between World Wars I and II, Kent, Dederick, McShane and others developed further applications of Statistics in connection with Ballistics. The staff of the Bell Telephone Laboratories, especially Dr. Walter A. Shewhart and Harold F. Dodge, was most fruitful in the discovery of Statistical techniques, and the Army was a shameless plagiarist in adapting them to its problems. Shewhart's work [6] led to the Army's first full-scale industrial use of Statistical Quality Control in manufacture at Picatinny Arsenal, Dover, New Jersey, which also was certainly one of the first few of such uses in the world. The Army Ammunition Surveillance [7] (Stockpile Reliability) System (circa 1939) was based largely on what was very recent work at that time. The Dodge-Romig Sampling Tables, not yet in book form [8], appeared just in time for use for ammunition inspection and acceptance tests in World War II. During the period shortly before World War II, the Army felt a bit smug about its statistical competence.

ARMY STATISTICAL PROGRESS DURING WORLD WAR II. World War II saw great progress in the military use of Statistics, due primarily to the availability to the war effort of men of competence. The National Defense Research Council (later, Office of Scientific Research and Development), the staff of the BRL, and, to a lesser extent, the staffs of Ordnance Arsenals acquired many Mathematicians and Statisticians of competence. Procedures for specifications of materiel, sampling, testing and interpretation of data (both planned data and the salvaging of unplanned data) were greatly improved. Indeed, Operations Research was being born even then. The Army* was not unmindful of the possible adaptation of any new Statistical "tool" to its work.

*References to the Army do not imply that the Navy and Air Force did not also make progress.

In addition to the above uses of Statistical Methods substantial progress was made by the Army during World War II of which there is little or no record. Many new techniques such as Sequential Sampling and Reliability were actually used in the Army, at least in an empirical way, before they were later designated by appropriate specific names. Of course, needed theory was not worked out in a formal way at that time. For example, the formal presentation of sequential sampling had to await the work of Dr. Abraham Wald, which was not published in book form until 1947 [9].

ARMY STATISTICAL PROGRESS, WORLD WAR II - 1953. After World War II, progress continued, although its rate was diminished due both to decrease in staff and to loss of some of the more competent people. Apparently, experiments that involved Factorial Designs were the first instances of full use of Experimental Designs in the Army. Factorial designs were used at the Ballistic Research Laboratories in the study of armor plate (1946-47)*, in the mammoth experiment on Aircraft Vulnerability (1946-50)*, and even on Project Stalk (a tank-fire control study under field conditions)* circa 1953. In 1953-1954 Reliability [10], in its present day sense, was used by Ordnance Research and Development, in a full-scale organizational and technical way, as a means of rescuing the Country's first operational guided missile, the NIKE, from a serious threat of failure.

With this rather glowing account of Army progress and status, one might well question wherein was the Army laggard, and where was the failure or potential threat of failure? What great work was there left to be done by the series of conferences on design of experiments under Wilks? I shall show that a very great deal was wrong with the Army's use (or lack of use) of statistical methods; that the task of righting the wrong was formidable, both in magnitude and in potential obstacles; and that astonishing progress has been made on the task during the nine years of the conferences.

From the survey of the 30 Army facilities, Wilks must have understood rather well what the Army needed, and have understood also the need for newly organized and sustained effort to supply the need. His skill as a teacher must have fortified him from fear of failure in undertaking to change the mode of operation of a large segment of the Army.

*Ballistic Research Laboratories Publications.

WHAT WAS WRONG. Let us observe that the origin, growth, and use of Statistical Methods in the Army was not only unplanned, but actually tended to progress in the least advantageous direction; i.e., from end-point to origin, rather than from origin to end. Roughly speaking, we can regard the military regime as consisting of the following steps or stages: doctrine, tactics, organization, selection of equipment, fabrication of equipment, test of equipment, and use of equipment. Logically, a powerful medium for the improvement of a stage should be first applied to the preceding stage or stages to which it is applicable. For example, a big improvement in use of equipment, (e.g., accuracy of ammunition) loses much of its potentially beneficial effect if either the tactics, organization, or weapons system is poor.

Contrary to the above observation, the earliest use of probability theory by the Army was for use of equipment, viz, the adjustment of artillery fire. The use of techniques based on the Gaussian Distribution, or Normal Probability Law, in connection with artillery fire probably is exceeded in antiquity only by the use of elementary probability theory in connection with games of change [11].

Decades elapsed before the next major step. In 1936, the Army began to use Statistical Quality Control in the manufacture of equipment, viz, the production of ammunition at Picatinny Arsenal, Dover, New Jersey. Kindred techniques such as sampling theory and statistical methods for analyzing data soon spread to improve specifications, inspections and acceptance tests.

During World War II almost all fabrication of military equipment was better, cheaper, and quicker, due largely to these techniques. During World War II, one strange reversal occurred in the inverse order of progress. Operations Research was born out of military sponsorship and was actually used to a limited degree by the staffs of high military planners in connection with the planning of the operations of large combat forces.

After World War II, it began to be more and more realized that since Statistical Methods improved the quality of equipment and reduced costs it would be a good idea to use similar techniques with the research, developing and testing in connection with new designs of equipment, thereby making better and more useful equipment designs at the out-set. Except for the invention of Reliability, which was a distinct child of necessity, this is just about where Wilks came in.

WILKS' TASK. When Wilks toured the 30 Army installations with the Army Mathematics Advisory Panel, it was he who articulated, "the most frequently mentioned needs expressed by the scientific personnel were for greater knowledge of modern statistical theory of the design and analyses of experiments." Thus, it is clear that Wilks recognized at least a major part of what was wrong with the Army; i.e., insufficient use of Design of Experiments in Research, Development and Testing.*

Certainly Wilks was not the first person to recognize the fact that an improvement in the early stages of the Army regime, i.e., doctrine, tactics, organization, etc., has greater leverage power than an improvement in later stages such as selection of equipment, fabrication, and use. The trend toward "up-stream" improvement began long before he appeared on the scene; and ranged from such measures as advocacy of industrial preparedness, as an important measure towards preserving the peace, to various stratagems for introducing sophistication in the upper stages of the Army's evolutionary process. Many persons deplored the fact that traditionally we had been forced to begin wars with the weapons left over from the previous war. Army Ordnance began to take measures against this ill shortly after World War I, and the then infant Army Ordnance Association (now the American Ordnance Association) lent a patriotic and helping hand, pursuant to its slogan advocating industrial preparedness as an insurance against war; i.e., a large production capacity should exist to meet a war demand for munitions of the latest designs. Army Ordnance realized that it must have an eye to the future and an ear to the ground regarding the plans and needs of the combat soldier, and therefore sent selected Ordnance Officers to the Army Schools ranging from the Command and General Staff College to the National War College to give them a close understanding of the combat soldier. Liaison officers from the combat arms were assigned to Aberdeen Proving Ground, Maryland, to assist in the realization of combat viewpoints, and in the development tests of materiel. Shortly after World War II, a number of persons, including some Ordnance, advocated the establishment of a scientific staff at Headquarters, Army Field Forces, Fort Monroe, Virginia, to assist in analyzing Army needs and in stating needs for new materiel in valid form. Such a group was partially formed and existed for a year

*The Army was not new to Wilks. In 1948 he was awarded a Joint Army-Navy Certificate of Merit for his war-time contributions to anti-submarine warfare and the solution of convoy problems.

or two*. However, it was Wilks who undertook systematically the task of greatly accelerating the spread of powerful and useful statistical techniques to the upper echelons of the Army regime, where the improvements that they enhanced would have the greatest leverage power.

Even if Wilks recognized the full nature of the job that he was doing, certainly, he did not have opportunity to finish the job. Much remains to be done. The real point in this discourse is the breadth and extent of the progress made in the nine years of Wilks' kindly and sympathetic leadership, effective persuasion, and his engendering of mutual cooperation and helpfulness between men of competence with whom he dealt. Let us try to note the progress, before any attempt to assess the remaining task.

ASSESSING THE PROGRESS. I hope that by the foregoing discussion I have led no one to believe that I have an objective method of measuring the progress of use of statistical methods in the Army during the 1955-63 period. I might say that the measuring of progress in a field of science or engineering is perhaps one degree more difficult than measuring the quantity and quality of output of research by laboratory; and whereas many have tried to do this, I know of no one who has really succeeded. The cold statistical facts are briefly these:

All the design of experiments conferences were for three days each, held in October or November, and conducted at a number of Army R&D establishments.

The number of registrants or conferees was always of the order of 200. Attendance was by invitation and the number of invitations was undoubtedly conditioned by the available accommodations.

The number of papers presented at each conference was of the order of 30. This appears to be about the number of papers that can be presented in a three-day conference.

All conferences were of a three-part character: Invited papers by distinguished Statisticians, technical sessions in which there were discussions of recent accomplished work, and clinical sessions in which work in progress was discussed from the viewpoint of inviting advice and criticism.

---

*Later, a permanent group was formed.

It thus appears that based on documental evidence the progress of the conferences can be judged only by the kinds of scientific and technical fields covered by the papers and by the inherent quality of the papers.

CHARACTER OF PAPERS PRESENTED. By and large, the place at which the conference was held had a strong influence on the character of the papers presented. This is undoubtedly due to the fact that the program committee gave some degree of precedence to the host institution, e.g., more papers bearing on the field of medicine were presented at the Eighth conference held at Walter Reed Medical Center than at other conferences. However, in the statistical fields there was a constantly increasing emphasis over the the nine years on the more sophisticated phases of design of experiments, screening theory, simulation stratagems, reliability, and techniques for evaluation of experiments. It is thus apparent that expertise on the part of the participants increased and also evident that the use of statistical experts in various fields of Army activities was increased both in number of experts and in variety of fields of activity.

Whereas, at the beginning of the conferences papers centered largely around items of Ordnance materiel, as the conferences proceeded the subject matter of the conferences expanded to include more emphasis on systems analysis. Similarly, with the penetration of statistical methods into new fields of activity, more papers were devoted to other than Ordnance equipment. With the broader use of statistical designs, papers appeared on the relation of equipment to organization, and to new theoretical developments having immediate application in Army use.

A further change in the character of the papers is the noticeable effect of learning to do by doing. It is apparent that whereas designed experiments gave greatly improved results, the same experiments also showed deficiencies in understanding what one's work was really about. For example, biases in results could be detected that were readily attributable to repeated use of the same personnel over the same terrain. Command exercises had to be altered and new stratagems employed (such as randomization techniques) to screen out the biases which passed unnoticed when experiments were of less sophisticated character. In fact it was precisely the acquirement of such evidence that convinced even non-statisticians that there was need for more movement "up-stream". This was a very fortunate circumstance because it drew military commanders into participation in the planning of the experiments and resulted in a constant movement of the sphere of

activity of statisticians into the domain of persons who were concerned with policy, tactics, and doctrine. Thus, non-statisticians saw the gains made through experiments in which they, themselves participated.

It is quite one thing to make a presentation on the efficacy of a technique, and quite another thing to convince the hearer that the use of the technique is important to his job. Successful experiments in which one himself has participated (although a step-wise process) are an effective method convincing one of the value of the methods used. By way of contrast, I believe that it would be quite impossible to suddenly inject into the military service (or into any other organizational sphere, for that matter) the concept and attitude which is expressed by the following quotation taken from a Combat Developments Experimentation Center (CDCEC) pamphlet:

> "The ability of the Army to carry out its goals in the future depends upon the success it has in achieving its combat developments goals today . . . of developing future concepts, doctrine, tactics, and techniques, and providing requirements for weapons, equipment, and appropriate organizations."

It is indeed heartening to read such a quotation. This Experimentation Center has an area of over a quarter of a million acres, a brigade of troops, a contract with Stanford Research Institute for Statistical Support, a variety of sophisticated equipment, including facilities for computer simulation of field experiments. Nevertheless, we know well that the tasks expressed in the quotation are only beginning and that only the first fruits have yet been achieved. From the foregoing example of CDCEC we can infer (a) that the advance of Statistical Methods in the Army, during the past nine years have been great, and (b) that the remaining part of the task, i.e., achieving the full nature of the job that Wilks undertook is still a large one.

WILKS' METHODOLOGY. If we hope to carry on in substantial measure the task that lies ahead we should take a good look at Wilks' methods. Wilks was a scientist for the sake of science, but he was also a realist and wished to see the practical results of applied science come to full fruition.

This is a rare combination of qualities.* Despite his many high scientific achievements and the respect in which he was held by his colleagues, he never assumed an authoritative position. On no occasion did he attempt to do a whole job himself to the exclusion of others. On the contrary, he always invited the cooperation of every person who could contribute substantially to getting the job done. He could organize and delegate without being obvious about it. In this way he secured the enthusiastic support of the men around him. If anything, he was more the servant of others than one demanding services. He had confidence in himself, but he also inspired confidence in others that led them to venture to cooperate, to work with him and to work together; and the work became an interesting enterprise to the point of preoccupation. In closing, I would like to give a brief example of how the spirit of Sam Wilks worked towards getting things done whether they were large or small.

AN EXAMPLE OF WILKS' WORK. About a year and a half ago, a gentleman in Georgia, a former member of the war-time team at The Franklin Institute, who is intensely interested in small arms fire asked several statisticians including Wilks some questions about the inter-relations of various measurements of central tendency and dispersion of shots on small arms targets, although he did not express it in these terms. In order to answer his questions, one needed to know the probability density distributions of several statistical measures whose distributions were unknown. These questions set off a kind of chain reaction. It was possible that answers to the small arms problem could well be answers to other, and probably more important, problems. Scientific men of good will, infused by the spirit of cooperation and scientific inquiry contributed what they knew to the general problem; but it became evident that a complete answer could be achieved only by some research that would add a modicum of knowledge to our existing store. Perhaps the most important contributions came (later) from Wilks, Grubbs, and one or two other colleagues in connection with their work on the analysis of tracking data on firings of long range missiles at the

---

*In writing for the Journal of the Royal Statistical Society, July, 1964, the noted British Statistician, E. S. Pearson says, "... it is hard to think of any mathematical statistician of the past 30 years who combined to a greater extent an excellence in the field of theory with a power of inspiring confidence in government agencies, national research institutions, and educational authorities, as a wise counseller in practical affairs."

Atlantic Missile Range, The work turned out to be so important that it has been carefully written up by Grubbs in a forthcoming monograph. This illustrates the humbleness, the spirit and the methods of Wilks. First, he was willing to lend his powers to anything that appeared to be a valid scientific enterprise; second, he had a keen perception of what is fundamentally important even though the context in which it was presented made it appear somewhat of casual interest if not unimportant; third, he could engender the spirit of true scientific inquiry into his colleagues; fourth, he could bring a matter to a crux so as to make it a permanent addition to the useful knowledge of mankind.

THE WILKS' AWARD. It is important that the spirit of Sam Wilks be carried on, both for an unselfish reason and a selfish reason. Our first reason is that of honoring his memory in gratitude for what he had done for us. The second and selfish reason is that carrying on the spirit of his work will contribute much to advancing the solutions for the great task that he loved and to which he devoted himself. We shall never achieve the task in full; but each solution or partial solution will contribute to the improvement of the military posture and safety of our Country. I am sure that Sam would approve this second motive. Through the generosity of Mr. Philip G. Rust of Thomasville, Georgia, and the good offices of the American Statistical Association, it appears that a means has been found of achieving, at least in part, both of the above purposes. An award will be created which by its character will help to carry on the stimulus of Wilks to Army Statistics.

REFERENCES

1. BRL Report No. 897, Proceedings of the First Symposium on Statistical Methods; Sampling Techniques.

2. Summary Report on the Establishment of the Army Mathematics Center. 14 October 1955. Office of Ordnance Research (now Army Research Office - Durham).

3. A. A. Bennett, On the Accuracy of Sampling. 6 January 1918. Unpublished paper in the files of the Ballistic Research Laboratories.

4. Fowler, Gallip, Lock, and Richmond, Aerodynamics of a Spinning Shell, Philosophical Transactions of the Royal Society of London, 1920.

5. F. R. Moulton, New Methods in Exterior Ballistics. University of Chicago Press. 1920.

6. W. A. Shewhart, Economic Control of Quality of Manufactured Product. Van Nostrand, 1931.

7. The Proposed System of Surveillance of War Reserve Ammunition, Zornig and Simon, BRL Report No. 115, 1938.

8. Dodge, Harold F. and Harry G. Romig, "Sampling Inspection Tables -- Single and Double Sampling", John Wiley and Sons, New York 1944; 2nd edition, 1959.

9. Abraham Wald, Sequential Analysis. John Wiley and Sons, 1947.

10. L. E. Simon, The Relation of Engineering to Very High Reliability. Proceedings of the Tenth National Symposium on Reliability and Quality Control. January 1964. IEEE.

11. Games, Gods and Gambling, F. N. David, 1962, Hafner Publishing Company, New York.

THE WILKS AWARD

Introduction of Mr. Donald C. Riley by Major General Leslie E. Simon

Mr. Chairman, Fellow Conferees, Ladies and Gentlemen, what the next two speakers have to say is so closely associated with my discourse on Wilks that I have been designated to introduce them.

As I implied at the end of my talk, the establishment of the Wilks Award was a tri-partite undertaking: And involved the Army as principal beneficiary, The American Statistical Association as the bearer of the burden of administration, and Mr. Philip G. Rust who endowed the award. Secretary Hawkins has personally expressed to the ASA his gratitude for its competent and patriotic services.

Mr. Donald C. Riley, Secretary-Treasurer and Executive Director of the American Statistical Association has rendered invaluable assistance in getting swift solutions to procedural problems he has been so kind as to agree to indicate to you the duties and obligations of the ASA in carrying out the Wilks Award; and he will also announce the recipient of the initial Wilks award. Don Riley!

INITIAL WILKS AWARD PRESENTED TO DR. FRANK E. GRUBBS

Donald C. Riley, Executive Director,
American Statistical Association

Many members of the American Statistical Association, as well as I, are glad to be present at this, the Tenth Annual Design of Experiments Conference. This is a very special occasion and the American Statistical Association is glad to participate during a uniquely auspicious time in its long history. This year is the 125th Anniversary of the establishment of the American Statistical Association which recent research at Stanford has found to be the second oldest national professional society in the United States.

The American Statistical Association has always worked closely, although usually quite informally, with agencies of the Federal Government. For example, during the year it was founded, 1839, it began to press for the improvement of decennial censuses and its representatives played a major part in the design of four of the six census schedules for the 1850 Census. As statistics and statistical methodology proliferated vastly since that time,

almost all areas of research have felt their impact. Certainly the whole area of design of experiments has had the closest association with statistics. The annual Design of Experiments Conference has become an institution. General Simon has reminded you of the close association of Professor Samuel S. Wilks with this Conference. Most of you know that relationship by heart. Sam lent his aid readily, unstintingly and effectively in many areas. This was part of the genius of the man.

I should note also that Wilks was the President of the American Statistical Association in 1950 and that he had always done much for the Association. He also helped to carry on in another area the close relation between the Association and the Federal Government. Just the day before he died he participated as a member of the Advisory Committee on Statistical Policy to the Office of Statistical Standards in the Bureau of the Budget. The Office of Statistical Standards requires consultation from time to time at a high level in its work as the central statistical coordinating body of the Federal Government. This Advisory Committee consists largely of former ASA Presidents and Wilks was one of its "founding fathers."

As mentioned in General Simon's address, the ASA has recently had the opportunity to be of further service. By joint agreement between representatives of the Army, Mr. Philip G. Rust and the ASA, the Samuel S. Wilks Award has been established. The Award will consist of a medal and an honorarium. The ASA has accepted the obligation of administering the Award in accordance with guidance and criteria which are consonant with law and with the wishes of Army representatives, Mr. Rust and the ASA.

Annually, ASA has agreed that an appropriate committee be selected (or appointed) to select the awardee, based on the criterion that he is a person whom the committee regards as deserving of the award, based primarily on his contribution (either recent or past) to the advancement of scientific or technical knowledge, ingenious application of existing knowledge, or successful activity in the fostering of cooperative scientific efforts which have only coincidentally benefited the Army. The award shall be made with the intent of recognizing the personal and intellectual accomplishments of the individual and shall not be given with the intent of supplementing the individual's salary, providing him with compensation, or advancing the interests of the donor or trustee of the endowment.

The American Statistical Association has been asked to invest the funds so generously turned over to it for this purpose and I am sure that its Board

of Directors, which has given its wholehearted approval, feels honored in being asked to join in honoring Sam Wilks. ASA will need to consult very closely with those of you who have helped to develop the annual Design of Experiments Conferences, in the selection of an Annual Sam Wilks Award Committee. I believe that Dr. Albert H. Bowker, the President of the American Statistical Association this year, will be able to announce this Committee shortly.

As executive Director of the ASA, I have the honor to announce that Dr. Frank E. Grubbs of the Army's Ballistic Research Laboratories has been selected to receive the "initial," not the first, Samuel S. Wilks Award. As is not unusual in the initial award of an honor, Dr. Grubbs was selected not by the process governing the first and subsequent recipients, but rather by unanimous agreement of those concerned with the establishment of the Award. He is so selected because of his close working relationship with Wilks, and especially because of his contributions along with Wilks to solutions and clarification of simple measures of dispersion, which are deemed useful to riflemen, ballisticians, and statisticians in general.

I have no medal to present to Dr. Grubbs, because the medal has not yet been struck; but it will be presented at the earliest appropriate opportunity, after it is available.

Incidentally, I will not be able to attend the banquet here tomorrow evening because I agreed long ago to attend the inauguration ceremonies in New York of Dr. Bowker as Chancellor of the combined Universities of the City of New York which was organized a few years ago.

The American Statistical Association will want to continue to advise closely with the Conference and will be glad to ask its auditor to render a brief auditing report each year if this seems satisfactory to those who have been so close to Sam Wilks, General Simon and especially Mr. Philip G. Rust, who has been so generous and public spirited in making the award possible. I should like to join in thanking Mr. Rust most profoundly.

INTRODUCTION OF MR. PHILIP G. RUST BY MAJOR GENERAL LESLIE E. SIMON

Mr. Chairman, Fellow Conferees and Ladies and Gentlemen.

We now come to the third and last speaker in this phase of our honoring Sam Wilks, Mr. Philip G. Rust of Winnstead Plantation, Thomasville, Georgia.

Mr. Rust is a very modest man, and more adept at understatement than a typical Britisher. It was only under pressure personally exerted by Secretary Hawkins that we succeeded, first, in overcoming his insistence that he remain anonymous, second, in getting him to attend this conference, and third, in persuading him to present the honorarium to the initial recipient of the Wilks Award, Dr. Grubbs.

Mr. Rust purports to be practically innocent of theoretical and applied statistics; but if under pressure, he can cite statistical literature by page and paragraph showing each historical advance in statistical measures of dispersion; he professes no close association with science and engineering, but I find that he was not only a research chemist for over ten years, but also returned to science and engineering during World War II; he lays claim only to being a Georgia farmer, but he has contributed to ASA the funds necessary to establish the award commemorating his old friend, Sam Wilks, contributing to the welfare of the military services, and fostering science in general.

With these cautionary remarks, I deem it a privilege and an honor to introduce Mr. Philip G. Rust.

THE CONCEPTION OF THE WILKS AWARD

Philip G. Rust
Winnstead Plantation, Thomasville, Georgia

Mr. Chairman and members of the audience you have heard a most informative talk by General Simon on "The Stimulus of S. S. Wilks to Army Statistics." Then, on Thursday we may look forward to Dr. Eisenhart's "Sam Wilks as I Remember Him."

In view of the newly established Association's Wilks Award, concisely described to you by Mr. Donald C. Riley, the Executive Director of the American Statistical Association; it is appropriate that I briefly discuss the conception of this award.

Back in the dark days of 1944, Dr. Wilks and I were headed north from Washington, by train; he to Princeton, and I to my home in Wilmington. At the time, I was at The Franklin Institute, working on .50 calibre barrel erosion, and also as the un-official translator of pertinent technical works. In passing, I would state that the Institute work was less statistical than of the ear drum rupturing variety.

On this train trip, I happened to mention, that for years, my spare time had been devoted to certain statistical measures of shots on a target. After telling Dr. Wilks about the firing of hundreds of .22 calibre targets, from rest; to get an empirical measure of the distribution of "extreme spread", he asked if I had started any theoretical work on the subject. (Incidentially, "extreme spread" is defined as the separation distance of the two widest apart shots.) His interest increased when it was mentioned that I had made a start by generating a few hundred artificial targets by using pairs of random numbers in the well-known bi-variate circular distribution. Equal likelihood of angular distribution was assumed, with no systematic errors.

The shots were laboriously plotted on cross section paper, and the extreme spread and other parameters examined. It is of interest to note that the fired targets and the plotted ones are extremely close.

About this time, my travelling companion suggested that he disembark at Wilmington, also. I had the feeling that he wanted to explore the application of these data to other, more vital matters. He stated that he had an exceptional graduate student who might be given the job of finding the true distribution of "extreme spread".

Eight or ten years went by, and our contacts were largely by phone. He assured me that he was still interested, and working on target problems; but that as yet, this distribution had not been discovered. The possibility of Monte Carlo methods on a to-be-aquired computer were discussed. Then on 10 August 1963, I received a long-hand letter saying that a 7090 computer was at hand, busily working on related matters.

While waiting for promised data from Dr. Wilks, I approached General Simon about the subject. He later discussed it with Dr. Frank Grubbs of Aberdeen, who subsequently brought forth an extremely useful manuscript, soon to be published.

Finally, on Dr. Wilk's 1963 Christmas card, he stated that the target problem was tied in with tracking work on the Atlantic Missile Range.

General Simon; with his very orderly mind, and sense of the fitting, then suggested the idea of the annual A. S. A. Wilks Award. This idea was greeted enthusiastically by all concerned.

What, then could be more fitting, than that Dr. Frank E. Grubbs should be the recipient of the initial award.

And now, it gives me great pleasure to hand Dr. Grubbs the initial honorarium and the assurance of its accompanying medal on its completion.

DEVELOPMENT OF THE DESIGN OF EXPERIMENTS OVER THE PAST TEN YEARS*

Oscar Kempthorne
Iowa State University, Ames, Iowa

INTRODUCTION. The main aspects of experimentation on which progress has been made in the past 10 years appear to be the following:

(a) the analysis of experiments
(b) the development of incomplete block designs
and (c) the investigation of multifactorial systems.

I shall have just a few words to say about the *first* two items and shall spend practically all my time on the third item.

THE ANALYSIS OF EXPERIMENTS. In the last 15 years or so, statisticians have become concerned about the assumptions that are commonly made in the analysis of comparative experiments. The common analysis is to use the matrix model

$$y = X\beta + e$$

in which y is the vector of observations, X is a matrix of known elements, β is a vector of unknown parameters, and the vector e of errors is assumed to consist of components which are normally and independently distributed around zero with constant variance. The obvious questions about such a model are:

(1) why use y, and why not a defined function of y, such as $\log y$ or $1/y$, or any of a host of other possibilities?

(2) is the model linear in the parameters, that is, is the expectation $X\beta$, correct?

(3) is the assumption about the errors correct?

In recent years there has been considerable attention to these questions, primarily by Anscombe (1961), Tukey (1962) and Anscombe and Tukey (1963), the work dating back to Tukey's one degree of freedom for non-additivity. This has led to the topic - residual analysis - which is now an every day phrase.

*Prepared in connection with work on Contract AF 33(615)-1737, Office of Aerospace Research, United States Air Force, Wright-Patterson Air Force Base, Ohio.

Matters unrelated to residual analysis but part of data analysis, are topics such as the question of multiple comparisons, the effects of preliminary test on conclusions, random, mixed and fixed models, and randomization theory of experimental inference. I shall not discuss these.

THE DEVELOPMENT ON INCOMPLETE BLOCK DESIGNS. Incomplete block designs were developed to control variability among the experimental units. The original incomplete block designs were given by Yates in the 30's, and in 1939 Bose and Nair developed a fairly general class of such design. Since that time there has been a development of blocking theory with regard to

(a) The structure and existence of incomplete block plans
(b) the arrangement of factorial designs in incomplete block designs.

Such development is very desirable, but it is agreed by most, I imagine, that the impact of this work on the conduct of experiments is not great. Roughly speaking we have had for many years an array of incomplete block designs which provides an adequate basis for choice for most experimental situations.

THE INVESTIGATION OF QUALITATIVE FACTORIAL SITUATIONS. It is essential to differentiate between multifactorial situations in which the factors are qualitative and in which the factors are continuous or quantitative. In the former case the structure of the totality of possible information consists of the true yields and variability for each of the possible factor combinations. In the latter case the totality of possible information is a functional relationship of yield to the levels of the factors or variables. So in the qualitative case, if one has factors say $a, b, c, \ldots,$ with levels denoted by $a_i, b_j, c_k, \ldots,$ the underlying formula for yield will be of the form $y(a_i, b_j, c_k, \ldots,) =$ $f(i, j, k, \ldots,)$ + error where the function is defined only for the factor levels $i, j, k, \ldots,$ in the situation. In other words, the model has to be a classificatory model. Classificatory models can be linear as exemplified by

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \text{etc} + \text{error},$$

or can be non-linear, as for example

$$y_{ijk} = \frac{\alpha_i}{\beta_j + \gamma_k} + \text{error.}$$

Essentially no theory exists for non-linear classificatory models, and I am of the opinion that this is a real gap in our knowledge.

In the case of study of the full set of factorial combinations, one of the basic problems is error control and systems of confounding were developed for symmetrical systems in the 30's. There have been a few developments in recent years with regard to confounding for the asymmetrical case, and also some clarification of the mathematical structure of factorial experiments [e. g. Kurkjian and Zelen (1962)] . I imagine, however, that examination of the full set of factorial combinations is rarely appropriate except possibly

(a) when most of the factors have 2-levels, with perhaps two three-level factors,

and (b) in the case of experiments, like in agronomy, for which there is a long essentially unalterable interval of time from executing the design to obtaining the experimental results, on the basis of which to plan another experiment.

There has been one development of analysis which seems to be very informative, when the totality of treatment degrees of freedom can be partitioned into meaningful orthogonal single degrees of freedom, the half-normal plot of Daniel (1959). The idea of half-normal plotting is the very elementary one of looking at the distribution of the totality of single degree of freedom contrasts, and to observe which ones are outliers. The half-normal plot is a convenient way of doing this. In general tight rules of significance for examining the realized half-normal plot do not exist. The procedures of half-normal plotting have been generalized to the case of a multivariate response by Wilk and Gnanadesikan (1963, 1964).

In the case of the linear classificatory models it is obvious that the simplest design problem is to estimate the effects under the assumption of no interactions. Effective designs for this case have now been available for several years. The earliest example of such a design was given by Tippett and is described in Fisher's "Design of Experiments" for the testing 5 factors in 25 trials. In the 1940's the following sets of main effect plans were developed:

the Fisher series:

$$2^n - 1 \text{ factors at } 2 \text{ levels with } 2^n \text{ trials}$$

$$\frac{p^n - 1}{p - 1} \text{ factors at } p \text{ levels with } p^n \text{ trials}$$

the Plackett-Burman series (related to Mood's weighing designs)

$$4N - 1 \text{ factors at } 2 \text{ levels in } 4N \text{ trials.}$$

An additional series was developed by Addelman and Kempthorne (1961) for

$$2\frac{(p^n - 1)}{(p - 1)} - 1 \quad \text{factors at } p \text{ levels in } 2p^n \text{ trials.}$$

In all these cases p is a prime number or a power of a prime number. Tukey (1959) and Addelman (1962) showed that these symmetrical main effect plans can be used to develop very reasonable main effect plans for asymmetrical factorial situations.

In the 1940's Finney (1945) formulated the general idea of fractional replication, which is closely related to the idea of confounding. It is interesting to note, in passing that Fisher was primarily interested in systems of confounding, and it was not adequately realized for some years that he had in fact developed incidentally the series of main effect plans mentioned above. The idea of fractional replication is to use a subset of the totality of treatment combinations chosen on the basis of the definition of effects and interactions. Obvious candidates as useful designs in this class are the main effect plans, and the designs which permit estimation of all main effects and two-factor interactions.

Also in 1946, Rao (1947) formulated the idea of orthogonal arrays. An array (N, k, s, t) is a collection of N treatment combinations out of the totality s^k of treatment combinations possible with k factors each at s levels, such that every combination of every subset of t factors occurs equally frequently. The value t is called the strength of the array. An array of strength 2 is an orthogonal main effect plan. With an array of strength 3, no main effect is confounded with two-factor interactions, but

some two-factor interactions are mutually confounded. An array of strength 4 enables the orthogonal estimation of all main effects and two-factor interactions, and so on. Clearly the enumeration of main effect plans, two-factor interaction plans etc. is related to the enumeration of orthogonal arrays. Box and Hunter (1961a, b) have given a rather detailed account of the possibilities of fractional replication with 2-level factors, using the term degree of resolution instead of the strength of array of Rao. A design of resolution III gives main effects estimates, which will be biassed by two-factor interactions. A design of resolution IV gives main effects unconfounded with two-factor interactions, but with the two-factor interactions somewhat interconfounded and a design of resolution V is a two-factor interaction - clear design. They show that a design of resolution III repeated with reversed signs gives a design of resolution IV. They discuss extensively the arrangement of fractionally replicated plans in blocks. They also examine the possibility of plans which estimate interactions among all of a subset of the factors with the effects of another subset of factors, the former being regarded as major variables and the latter as minor variables. For example, they give a 2^{16-11} plan which enables the estimation of all effects and interactions among 4 major variables and the main effects of 16 minor variables. Box and Hunter (1961b) give the possible two-factor interaction clear fractions in blocks for up to 11 factors. The possibilities are as follows:

No. of factors	5	6	7	8	9	10	11
No. of observations	16	32	64	64	128	128	128
No. of blocks	1	2	8	4	8	8	8 .

Addelman (private communication) has found a 2^{17-9} resolution V plan in 8 blocks of 32. These plans enable orthogonal estimation of all the effects and two-factor interactions and appear to be the minimal designs which allow orthogonal estimates.

If one is prepared to relax the orthogonality requirement, one can obtain reasonably precise estimates with irregular fractions (Addelman, 1961 and Whitwell and Morbey, 1961). For instance Addelman gives a fraction $\frac{3}{8}$ of a 2^7 factorial, $\frac{3}{16}$ of a 2^8, and $\frac{3}{16}$ of a 2^9 to estimate all main effects and 2-factor interactions. Whitwell and Morbey give a design using 96 observations which allows the estimation of the main effects and all but 3 of the two-factor interactions of 11 factors.

Fractional replication of the 3^n factorial system is much more difficult, as soon as one wishes to estimate two-factor interactions. In the case of 5 factors, for instance, the smallest plan which allows estimation of two-factor interactions is a 1/3 replicate requiring 81 observations. The problems of enumerating two-factor interaction clear plans for the 3^n factorial system appear to be rather difficult. Bose, Bush, Seiden and others have worked on the enumeration of orthogonal arrays and on the maximum number of factors which can be accommodated with a given number of observations, but the situation is still quite unclear. Obviously, the main experimental interest is in arrays of strength 4.

One possible way of examining a multifactor situation is by some use of random sampling of the totality of treatment combinations. This idea was first put forward, it appears, by Satterthwaite (1959) and attempts have been made to develop a theory of inference from such sampling, e.g. by Dempster (1960, 1961). It appears that the situation is very difficult. Ehrenfeld and Zacks (1961), Zacks (1963) and Ehrenfeld and Zacks (1963) have examined two procedures of random sampling the totality of treatment combinations which are based on fractional replication. It would appear that considerable further development is needed of ways of sampling the totality of treatment combinations and of analyzing the resultant sample.

The general moral to be drawn, then, with regard to multifactor (qualitative) experiments, is that it is easy to examine for main effects, more or less regardless of the number of levels, but that examination for interactions can in general be done at all easily only with two levels for each factor. It is likely that if the requirement of orthogonality is waived, plans requiring reasonable numbers of observations can be developed.

THE INVESTIGATION OF DEPENDENCE OF A YIELD VARIABLE (y) ON k CONTINUOUS CONTROL VARIABLES ($x_1, x_2, \ldots, x_k$). It would seem that while there are many aspects of the dependence of a yield variable on k control variables which can be varied continuously, one can "spin off" one problem which is quite different in nature from all the others, and that is the optimization problem, namely to determine the values of $x_1, x_2, \ldots, x_k$, such that the yield is a maximum (or minimum). Of course there are situations in which there are several yield variables, say, $y_1, y_2, \ldots, y_m$ and the problem may be more complex, such as to determine the combination $(x_1, x_2, \ldots, x_k)$ for which y_1 is a maximum, subject to restraints of the type $y_2 < k_2$, $y_3 > k_3$ and so on.

OPTIMUM SEEKING. The work in this area was utterly naive, one factor at a time experimentation, until the work of Box and Wilson (1951) to whom great credit is due for tackling the problem with some degree of sophistication. I shall enumerate briefly the steps of the Box-Wilson procedure. They are:

(1) local exploration around a guessed optimum by means of a design which enables the fitting of the relationship

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k ;$$

(2) proceeding along a line in the direction of steepest ascent in the units chosen to an optimum on that line;

(3) local exploration around this newly obtained optimum as in (1) ;

(4) proceeding along a new steepest ascent direction as in (2);

(5) repetition of steps (3) and (4);

(6) when there ceases to be a pay-off from this process, perform local experimentation around the achieved sub-optimum to enable the fitting of a second degree dependence of y on the x's;

(7) do a mathematical analysis of the achieved second degree relationship. That is, if one has found the relationship

$$y = \beta_0 + \Sigma\beta_i x_i + \sum_{i \leq j} \beta_{ij}x_ix_j ,$$

then one can make a linear transformation of $x_1, x_2, \ldots, x_k$ to say $z_1, z_2, \ldots, z_k$ so that

$$y = \gamma_0 + \lambda_1 z_1^2 + \lambda_2 z_2^2 + \ldots + \lambda_k z_k^2 ;$$

(8) this representation enables one to see the form of the relationship of y to the z's in the neighborhood of the sub-optimum achieved earlier. If all the λ_i are negative, the optimum is at the point where all the z's are zero. If some are zero there is a subspace of optima. If for example λ_1 is zero and the others are negative

the optimum (maximum) is achieved wherever z_1 which is a linear function of the x's is zero. If of course any λ_i is positive the maximum is not at all defined by the fit.

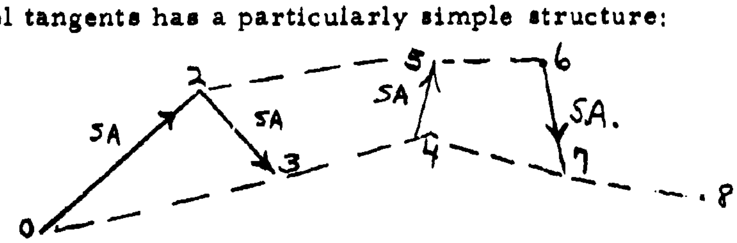
Apart from steps (6), (7) and (8) this is the standard iterated steepest ascent. Obviously the procedure was developed for the optimization of a production process in which only local experimentation is possible so as not to disrupt production.

The procedure suffers from the well-known disadvantage of steepest ascent in that progress may be excellent for the first few steps but then becomes very slow. Of course steps (6), (7) and (8) were inserted by Box and Wilson to take care of this.

A line of attack on this problem, which is closely related to the Box-Wilson approach, consists of trying to develop algorithms which will give rapid convergence to the optimum if the variable to be optimized y, say, is known without error and is of the form

$$y = b_0 + b'\alpha. + x'Cx$$

in which C is negative definite, so that a unique optimum exists. One then attempts to determine the properties of the algorithm if the relationship of the y to the x's is not of the postulated form, and if y is known only with error. The methods I know of which have this structure are the following, the method of parallel tangents due to Shah, Buehler and myself (1964), and the method of Fletcher and Powell (1963). The method of Fletcher and Powell is based on a guess of the matrix C, which would ordinarily be taken as the unit matrix, and on successive line searches, the directions of which change on the basis of previously determined gradients and on the steps to the optima on the lines. The method of parallel tangents is really just an acceleration of the initial steps of the Box-Wilson procedure which removes the necessity of fitting a second order relationship. One variant of the method of parallel tangents has a particularly simple structure:



in which the lines labelled S. A. are steepest ascent lines and the dashed lines are acceleration lines. In the absence of error and with k dimensional ellipsoidal yield contours the maximum as reached at the point labeled 2n.

There are other intuitive methods such as pattern search of Hooke and Jeeves (1962), and methods using sectioning of the factor space on the basis of tangent planes to the yield contours (Wilde, 1964).

These methods appear to use with some degree of effectiveness, the information that is accumulated by the separate local experiments. A real difficulty from a theoretical viewpoint is to evaluate the properties of all these methods, including the Box-Wilson method, in the presence of error.

Just how important it is from a practical viewpoint to establish tight clean mathematical results about the performance of these strategies in the presence of error is, I believe, a moot point. It would of course be valuable from an aesthetic viewpoint to have such information, but the difficulties of obtaining information of practical value seem to be tremendous. It is clear that the strategies described above are so loosely defined that they cannot be subjected to precise mathematical evaluation. Answers to such questions as (a) how does one explore locally? (b) what is the "spread" of the local design? (c) how does one search for the optimum on a line? (d) how does one decide when to terminate?, are not given by th procedures. They are, however, questions which the user will be able to make choices which must, of course, be somewhat arbitrary but which will be modified as information accumulates. If the local experimentation does not indicate clearly that there is a direction in which improvement can be made, more local experimentation will be done, presumably by either repeating what was done before or by "pulling in" the local design and repeating. Also, it is obvious that the experimenter will survey the totality of information obtained up to any particular point in the process and will modify the algorithms if he can spot a pattern in the response relationship.

A direct attack on the optimization problem with error was made by Kiefer and Wolfowitz (1952) with work related to that of Robbins and Monro (1951) who developed a stochastic approximation scheme for finding the value x, at which the expected value M(x) of a random variable y(x) takes a particular value. The Kiefer-Wolfowitz procedure is as follows: for the case of optimization in one dimension choose two sequences of positive numbers, c_n, a_n, such that $\lim c_n = 0$, $\Sigma a_n = \infty$, $\Sigma a_n c_n < \infty$ and $\Sigma a_n^2 c_n^{-2} < \infty$, as, for example $a_n = \frac{1}{n}$, $c_n = \frac{1}{n^{1/3}}$; take an arbitrary z_1 and then use

$$z_{n+1} = z_n + a_n \cdot \left\{ \frac{y(z_n + c_n) - y(z_n - c_n)}{c_n} \right\}.$$

Then z_n converges stochastically to the point z at which $E\, y(z)$ is a maximum. Kiefer and Wolfowitz (1952) state that there remain the problems of choices of sequences a_n and c_n which will be optimal in some sense, and the specification of a stopping rule. This line of work has been developed considerably by Blum (1954), Dvoretsky (1956), Kesten (1958) and by Sacks (1958), and others to the multidimensional case.

It is not clear at all what the attitude of the practical statistician should be to these very different approaches. Kiefer (1959) states that methods such as the Box-Wilson one or the others of the same flavor, "cannot in their present state have any role in satisfactorily solving these problems, since they have no guaranteed probability properties and are not even well-defined rules of operation." Barnard, however, in discussion of Kiefer's paper, disagreed and took the view that rules of operation which are not well-defined may be preferable to the rules which are. It would seem that the guaranteed property of convergence with probability one with an infinite number of observations is small comfort to the practical man, even though it was obviously not easy to develop procedures for which one can prove the property.

What we really lack are accounts of actual experiences with the various methods. Perhaps a good practical strategy is to use the "deterministic" schemes at first, and then turn to the stochastic schemes when the former cease to give advances.

RESPONSE SURFACE EXPLORATION. I now turn to the problem of studying the dependence of a yield variable y on continuous control variables $(x_1, x_2, \ldots, x_k)$ which has been termed a response surface exploration by Box and his co-workers.

The great bulk of the work on this problem has been by Box and his associates, stemming back to the famous Box-Wilson paper (1951). The background for the work is the paper by Box (1952) on first order multifactorial designs, which I have to review even though it was done more than

10 years ago. Here Box specified the amount of variation of each variable or factor by defining the scale unit S_i for the i-th variable as

$$S_i = \left\{ \sum_u \frac{(X_{iu} - \bar{X}_i)^2}{N} \right\}^{1/2} ,$$

where X_{iu} is the level of the i-th factor in the u-th observation. He defined the standardized variable x_{iu} as

$$x_{iu} = (X_{iu} - \bar{X}_i) / S_i \quad .$$

He then took the design problem to be as follows:

(a) the experimenter is to specify $\bar{X}_i$, the "center" of the design and scale multiplier S_i for each variable,

(b) the designer of the experiment is to choose an array of standardized levels, x_{iu}, at which the observations are to be taken, the actual levels being

$$x_{iu} = \bar{X}_i + x_{iu} S_i \quad .$$

In other words, the "center" of the design and the "spread" are specified by the experimenter and the only problem of the designer is to choose the x_{iu} which, of course, satisfy

$$\sum_{u=1}^{N} x_{iu} = 0, \qquad \sum_{u=1}^{N} x_{iu}^2 = N \quad .$$

I shall comment on this basis later, but, for the present, will indicate the subsequent developments. In the case of the first order designs, the criterion was optimum estimation of the coefficients in the equation

$$y_u = \beta_0 + \beta_1 x_{1u} + \beta_2 x_{2u} + \dots + \beta_k x_{ku}$$

and the optimum design is one in which the x_{iu} are given by the columns, after the first, of a matrix $N^{1/2}O$, where O is an orthogonal matrix whose first column consists of unit elements. Box then noted that if the number of observations is k+1, the experimental points are the vertices of a regular dimensional simplex. He also noted that any rotation of this regular figure would satisfy the conditions. Box and Hunter (1957) developed in considerable detail the concept of rotatability. A design is said to be rotatable if, when the levels of the variables are standardized as stated above to be $(x_1, x_2, \ldots, x_k)$, the variance of the predicted y at a point $(x_1, x_2, \ldots, x_k)$ is a function of these x's only through $\sum_1 x_i^2$. In other words if one were to construct contours of variance of the predicted y they would be spherical with center at the 'center' of the design, when plotted in standardized levels. They stated their aim to be "to develop arrangements which generate information (equal to the reciprocal of the variance of prediction of y) symmetrically in those coordinates regarded as most relevant to the experimenter." Box and Hunter developed second order designs in 2 dimensions by taking two or more concentric rings of points, with each ring being a regular figure, for example a pentagonal design with extra center points. For 3 dimensions, they took points equally spaced on a sphere, for instance, by combining a regular tetrahedron, a octahedron, and a cube with additional center points. For more than three dimensions they suggested the combination of the points of a 2^k factorial, and 2^k points of an axial set and additional center points. Throughout attention was paid to the problem of blocking, that is, of arranging the totality of points in subsets to enable the elimination of heterogeneity between the units. Box and Behnken (1960a) developed designs by operating in a simple way on first order simplex designs. If the points of the simplex design are regarded as vectors, one can develop additional points by forming sums of the original vectors two at a time, sums of the original vectors three at a time, and so on. The configurations so developed are then scaled to satisfy the scaling and rotatability conditions. In this way they obtained, for instance, designs to examine 4 variables in two blocks of 22 observations, 5 variables in two blocks of 26 observations, 6 variables in two blocks of 34 observations, 7 variables in two blocks of 33 observations. The last one in this list is quite impressive in that it uses only 3 levels of each factor and enables all 36 coefficients of a second degree fitting to be evaluated reasonably. It is curious that all the points except the center points be on a hypersphere of radius $\sqrt{3}$ (in the standardized units). Box and Behnken (1960b) developed

another series of 3-level rotatable designs by utilizing incomplete block configurations. The simplest example was the following. We have the balanced incomplete block configuration

		x_1	x_2	x_3	x_4
"Block"	1	x	x		
	2			x	x
	3	x			x
	4		x	x	
	5	x		x	
	6		x		x

If a "block" contains x_i and x_j, it is replaced by the 4 treatment combinations on x_i and x_j, (-1,-1), (-1,1), (1,-1) and (1,1), the other variables being taken at the zero level. Bose and Draper (1959), Draper (1960a) and others have constructed classes of second order rotatable designs. Gardner, Grandage and Hader (1959) and Draper (1960b, 1961, 1962) have developed third order rotatable designs. Throughout it appears that the designs are based on the combination of symmetrically placed points on spheres in the standardized factor space. The ideas of Box have led to the development of a considerable array of designs, all based on the concept of rotatability. Many of the designs are remarkable in that they allow the fitting of functions of the second or third degree with relatively low redundancy of experimental points. Also by choosing odd moments up to particular order equal to zero, one can prevent bias in the regression coefficients from third order coefficients in the polynomial representation.

The motivation for the development of the array of rotatable designs seems to be summarized by Box and Behnken (1960a, page 840) in the following quotation,

> "At a particular stage we are interested in the behavior of the response function 'in the neighborhood' R of some particular point P. We have in mind that the operability region O, that is the region in the space of the variables in which experiments could be conducted, is fairly extensive and that P is not close

to the boundary of O. We suppose that the neighborhood of interest about P is a region R which nowhere reaches the boundary of O and that scales, metrics and transformations are chosen either implicitly or explicitly such that R is very approximately spherical and is centered at P."

Essentially all the designs whose development I have mentioned earlier were aimed at controlling the variance of the prediction based on the fitting of a polynomial of the first second or third degree. There had been some attention to the bias in estimated polynomial coefficients from higher polynomial terms that were ignored in the fitting. Box and Draper (1959) made a direct attack on the problem of bias, within the framework of previous developments. The situation considered was that a function $f(x_1, x_2, \ldots, x_k)$ is fitted, when the true functional dependency is $g(x_1, x_2, \ldots, x_k)$. The mean square error of a prediction consists of the variance plus the square of the bias. Box and Draper consider the average over a region of interest R in the $(x_1, x_2, \ldots, x_k)$ space of these two components, for the particular case when $f(x_1, x_2, \ldots, x_k)$ is linear and $g(x_1, x_2, \ldots, x_k)$ is quadratic. They conclude that the optimal design is very nearly that which would be obtained if variance is ignored and only bias is considered. If this conclusion is accepted, it would appear that the whole class of rotatable designs based on variance considerations, need careful re-examination from the viewpoint of bias. The development depends strongly, it would appear, on the choice of the region of interest as being spherical in the standardized variables, and on equal weighting over the interior of the "sphere" of interest. The reasons for choosing this framework appear to be mathematical, in that with this framework, integrals can be evaluated. Box and Draper prove a theorem that is highly indicative of the nature of the problem. The theorem states that if a polynomial of degree d_1 is fitted by least squares over any region of interest R in the k variables, when the true function is of degree d_2, greater than d_1, then the average squared bias over R is minimized by making the moments of order up to $d_1 + d_2$ equal to the corresponding moments of a uniform distribution over R. So if one knew nothing about the true function except that it can be represented by a polynomial of indefinitely large degree one should spread the observations evenly over the region R. Clearly the definition of the region R should be made in terms of variables for which one could hope that a low degree polynomial would give a good fit.

The whole line of development appears, however, to suffer from some defects which are illustrated by the simplest designs that were developed -- the simplex first order designs. For the case of 3 variables with 4 observations, Box exhibited two designs which he claims to be equally good:

$$D_a = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \end{array} \qquad D_b = \begin{array}{c} \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{bmatrix} -\sqrt{2} & \frac{-\sqrt{2}}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \\ \sqrt{2} & \frac{-\sqrt{2}}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \\ 0 & 2\frac{\sqrt{2}}{\sqrt{3}} & \frac{-1}{\sqrt{3}} \\ 0 & 0 & \frac{3}{\sqrt{3}} \end{bmatrix} \end{array}$$

with

$$D_a'D_a = D_b'D_b = 4I$$

where I is the 3×3 identity matrix. These two designs have the same center and have equal spread with the definition of Box. However, if design D_b can be used, x_1 can be varied between $-\sqrt{2}$ and $\sqrt{2}$, x_2 can be as large as $2\frac{\sqrt{2}}{3}$, and x_3 can be as large as $3/\sqrt{3}$, whereas in design D_a the limits for each x are from -1 to +1. If however, the situation is such that one can vary the x's over the ranges specified in design D_b, one would be foolish in not varying them over the same range, with the first order design D_a, and if one does, the resultant design D_a^*, say, is clearly better as a first order design than the design D_b. The same criticism has been made by Kiefer (1961b).

This simple example brings to light one of the basic problems of exploration, as opposed to optimum seeking, namely, that the region of possible experimentation must be defined if one is to attempt to develop

a good design. The simple example above shows that the standardization of variables in terms of root mean square deviation of levels results in peculiar restrictions. It would seem more natural and appropriate to define the region of possible experimentation in terms of the original unstandardized variables. If one is exploring the relationship of a yield variable y to a single control variable X, a natural restriction would be that one can experiment at X values in a prechosen interval of X, say from $X = a$ to $X = b$. If one has two control variables X_1 and X_2, a possible specification of the region of permissible experimentation would be X_1 in the interval (a_1, b_1), and X_2 in the interval (a_2, b_2). It is, of course, quite likely that as soon as one has more than one variable, the region of possible experimentation will not be rectangular in the variables originally thought of. It is inconceivable that one will be able to develop a useful theory of experimentation for an arbitrary region of possible experimentation. It does, however, seem reasonable that one can choose "new" control variables that are functions of the originally thought of variables so that the region of possible experimentation in the "new" variables is approximately either a hypercube or a hypersphere. At least in this way one can set up a mathematically defined problem for which one can hope to get an answer. One might hazard the guess with the emphasis on sphericity that results from considerations of rotatability, that the rotatable designs will prove to be good designs in the case when the region of possible experimentation can be defined to be spherical. Some problems of allocation for polynomial regression within a spherical region have been considered by Kiefer (1961b) and are discussed below. It appears that a few of the Box-Hunter rotatable designs of very specialized nature are optimal with respect to two of the possible criteria. However the implications of the scaling in the Box-Hunter rotatable designs are obscure.

It appears, then, that a more fundamental approach to the problem of design would take as its base a definition of region of possible experimentation, provided by the experimenter. It is then necessary to formulate the aims of the experiment, and it is at this point that one opens a Pandora's box, because of the multiplicity of partially conflicting aims that always occurs.

Since the beginning of the formal development of designs there has been some attention to optimality of design. In the simple case of linear regression on an interval it has been known for decades that the best disposition of resources for estimation of the slope is to place half of the observations at each end of the interval. In the case of comparisons of two groups it is obvious that for maximum precision of the group difference one should have equal numbers of observations in the two groups. It is also obvious that if one has several groups, and one has the same interest in all possible differences of pairs of groups, one should, with homoscedasticity, have each group equally represented. Indeed the requirement of equal interest forces equality of representation. The classical symmetrical designs for error control, such as randomized blocks, Latin squares, balanced incomplete blocks, were considered good, because the prime interest of the experimenter was considered to be estimation, with equal interest in all the treatments, which were taken to be fixed. They were also based on the idea that the main difficulty of experimentation was to control variability between experimental units, and that variability within a group of experimental units was a monotonic function of group size.

Work on optimality of design was done early by Plackett and Burman who showed that the orthogonal 2^n plans or fractions of these, such as those based on Hadamard matrices were optimal in a useful sense for qualitative main effects of two-level factors. Indeed they resulted in as efficient estimation for each single parameter, as one could obtain if one used the whole of the experimental resources just to estimate that single parameter, and this, really, is much more than one was ever entitled to hope for. A few years later optimality of design was attacked frontally by Elfving (1952), Chernoff (1953) and Ehrenfeld (1955). The topic was taken up very extensively by Kiefer and Wolfowitz (1959) and Kiefer (1958, 1959, 1961a,b, 1962).

The whole problem of optimal design is of course, to decide what to optimize for. Kiefer (1959) lists several possibilities:

(a) maximizing the infimum of power of test of a null hypothesis against a class of alternatives (M-optimality),

(b) maximizing the limiting power of test in the neighborhood of the null hypothesis (L-optimality),

(c) minimizing generalized variance of estimates of parameters (D-optimality),

(d) minimizing the maximum eigenvalue of the variance-covariance matrix of estimates, used by Wald (1943) and Ehrenfeld (1955) (E-optimality),

(e) minimizing the trace of the variance-covariance matrix of estimates (A-optimality),

and

(f) minimizing the maximum variance of prediction over the experimental region (G-optimality).

These criteria can be applied to the totality of parameters or to a chosen subset of the parameters.

It needs to be emphasized, I think, that all these criteria are related to the problem of control of error with a model which is assumed to be true. It is not clear that designs which are good for error control are also good for detection of bias of model, as Box and Draper showed in work that I mentioned earlier. In the incomplete block problem, for instance, I am inclined to the view that designs which have some repetition of treatments within blocks are desirable. Such designs will be inefficient with regard to any of the above optimality criteria, if balanced incomplete block designs are possible, but will enable better examination of the adequacy of the usual additive model.

Kiefer (1958, 1959) has proved that balanced block designs, Latin squares, Youden squares, orthogonal arrays, are optimal with regard to criteria A, D, E and L. These results are, I suppose, of some mathematical interest, and suggest that if one has a balanced array of experimental units one should try to use the restrictions of the array. However they do not answer questions like whether one should use a Latin square design rather than a complete block design. The Latin square result states that if one is going to use the Latin square model for analysis one should use the Latin square design, and as such is not at all surprising.

Kiefer (1958, p. 676) characterizes M-optimality as "the strongest and least artificial of the four" criteria, D, E, M and L, and it was attention to testing of hypotheses that led Kiefer to give the examples which generated, apparently, much unnecessary heat at the Royal Statistical Society meeting. Kiefer pointed out that if one had 6 observations to be split among three populations which are $N(\theta_i, \sigma^2)$, $i = 1, 2, 3$, then different designs were optimal for the three problems:

(a) point estimation of θ_1, θ_2, θ_3

(b) testing the hypothesis $\theta_1 = \theta_2 = \theta_3 = 0$

(c) testing the hypothesis $\theta_1 = \theta_2 = \theta_3$,

where in (b) and (c) one is interested in alternatives near the null hypothesis. For problem (a) one should take 2 observations from each population, for problem (b) one should take one of the populations at random and use all 6 observations on it, while for problem (c), one should take two of the three possible populations at random and then take 3 observations from each. This example shows very clearly that different criteria of optimality can give radically different designs.

The work of Kiefer and Wolfowitz is more informative, I think, in the area of polynomial regression than in the area of qualitative experimentation. The history of optimum allocation for polynomial regression appears to be as follows. In the one-dimensional case for which the units can be chosen so that the interval of experimentation is $(-1, 1)$, Guest (1958) considered the G criterion above, the maximum variance of a prediction, and showed that this was minimized by placing $\frac{1}{k+1}$ of the points at each end of the interval and at the zeros of the derivative of the k-th degree Legendre polynomial. Hoel (1958) showed that if one wishes to minimise the generalized variance of the coefficients of a k-th degree polynomial the optimum allocation was the same as that obtained by Guest. Kiefer and Wolfowitz (1959) showed that the best estimate of the coefficient of x^h, when a polynomial of degree h was required for the x-interval $(-1, 1)$, was to place $1/2h$ of the observations at each end of the interval and $1/h$ at the points $\cos(j\pi/h)$, $1 \leq j \leq h-1$, which may be termed Chebychev spacing. In experimentation on the square $-1 \leq x_1 \leq 1$, $-1 \leq x_2 \leq 1$, the

best test of interaction term x_1x_2 is obtained by placing 1/4 of the observations at each corner. Of course all the above solutions depend on the total number of observations being appropriately divisible. Kiefer (1959) gives the example that with 4 observations, the best placement for cubic regression on the interval (-1, 1) is at the values ± 1, $\pm 1/\sqrt{5}$, and with 5 observations the best placement is at the values 0, $\pm$ 0.511, ± 1. The dependency of optimum design on the specific value of N is avoided by Kiefer and Wolfowitz who consider how best one would place an infinite number of observations. Such placements can be regarded as approximate designs, and they proved (1960) a rather remarkable theorem that the design using a large number of observations which minimizes the generalized variance of the coefficients of a polynomial fitting would also minimize the maximum variance of a predicted value over the experimental region. It is not clear just how useful this result is for reasonable numbers of observations, and how one should use the approximate placing given by the theorem to arrive at a placement for a reasonable number of observations.

With this proviso, however, this later work of Kiefer and Wolfowitz gives an indication for the choice of design in "response surface exploration," at least if one views the matter as a problem of polynomial approximation. The fact that the generalized variance of coefficients is minimized would tend to indicate (though it does not guarantee) that all the coefficients of a polynomial are being estimated with reasonable precision, and the fact that the maximum variance of a prediction is minimized should to a moderate extent permit the discovery of lack of fit by the polynomial.

In the case of quadratic regression on a hypercube bounded by -1 and 1 in each direction, in q(= 2, 3, 4, or 5) dimensions, Kiefer (1961) shows that the best "infinite" design is to assign a proportion α of the experimental points to each of the 2^q corners, a proportion β to the mid point of each of the $q2^{q-1}$ edges, and a proportion γ to the center of each of the $q(q-1)2^{q-3}$ 2-dimensional faces of the hypercube. In the case of q equals 5, the values of α, β, γ are

$$\alpha = .01928$$

$$\beta = .0003125$$

$$\gamma = .004475$$

However, in view of the fact that the α set contains 32 points, and the β and γ sets contain 80 points each, this "infinite resources" answer is not really useful. It does not tell us, for instance, how we should place say 50 or 60 observations. It does appear to indicate, however, that if the G criterion, which seems a somewhat superior one for exploration, is adopted, then the experimental points should be placed near the corners and edges of a rectangular experimental region. This is in considerable contrast to the rotatable designs discussed earlier, which seem to devote much attention to the center and interior of the region.

Later Kiefer (1961b) examined polynomial regression when the region of experimentation and interest is the hypersphere or "ball," $\Sigma x_i^2 \leq 1$. It might be expected that the designs he would get would be related to the rotatable designs in that the latter seem to be aimed at a spherical region of interest. Kiefer considers the approximate case, that is, the "infinite resources" case, so that D-optimality and G-optimality are equivalent. He was able to characterize partially the approximate optimal design, and showed that it is rotatable. In the case of linear polynomial fitting, the best design has equal weight at the vertices of an inscribed regular simplex or the vertices of any other inscribed regular polygon. So for this case the maximally spread simplex design of Box (1952) is optimum with these criteria. Also in two-dimensions with quadratic regression, the design with one observation at the center and one at each vertex of an inscribed regular pentagon is D-optimal and hence G-optimal. However, apparently most of the rotatable designs do not have these optimality properties. I cannot regard the lack of optimality properties as seriously as apparently Kiefer does. Kiefer (1961b, p. 398), feels that he justified for the first time the use of rotatable designs but I regard his results as mathematically rather elegant, and not totally relevant to the problems of the experimenter. The representation of yield as a polynomial in the control variables is unaesthetic and uneconomical of parameters, except in the optimization problem. Even in the optimization problem it is highly questionable whether one should do local experimentation other than to get gradients. I agree with Kiefer that the framework within which Box and his associates have worked has serious logical deficiencies, but also have the view that they developed some very useful designs and design ideas.

CONCLUSIONS ON THE EXPLORATION PROBLEM. The problem of studying the dependence of a yield variable on control variables is not well-defined. Experimenters with this problem will have a multiplicity of aims,

such as to obtain reasonably precise estimates, reasonable strength of evidence against particular null hypotheses of interest, ability to select a functional form that represents the data well and is economical of parameters, and so on.

The theoretical statistician can obtain optimal designs only by forcing the problem into a highly idealized simplified form, and there is a tendency to regard the optimal design for idealized simplified form as the design the experimenter should use. This attitude seems to be exemplified by Kiefer's remark (1959, p. 316), "Why not think in terms of the right space of decisions from the outset?" I have yet to meet an experimenter whose aims can be represented by a space of decisions, which is sufficiently well-defined to be susceptible to such an attack.

The work of the optimizers is, however, valuable, because it gives us suggestions of respects in which a plan may be weak. The upshot for me of the work I have reviewed is exemplified by the following cases. In the case of 3 factors in a cubic region (-1, 1), I would do the following:

(i) with 4 observations I would take a 2^{3-1} factorial at the corners;

(ii) with 9 observations I would use a 1/3 replicate of the 3^3 with levels -1, 0 and 1 for each factor;

(iii) with 15 points I would use the corners and center of each face and the center which is essentially a central composite design but not rotatable;

(iv) with 27 points I would use the full 3^3 factorial with levels -1, 0, and 1.

If the problem is really one of studying the dependence I would try to persuade the experimenter to do the full factorial (iv), because it would enable me to think, to some advantage, about representations other than by a polynomial. In the case of 4 factors, I would think with a low number of possible observations in terms of main effect plans with observations at the corners. If more were possible I would consider the sets of points:

S_1 : $(\pm 1, \pm 1, \pm 1, \pm 1)$

S_2 : $(\pm 1, \pm 1, \pm 1, 0)$ with permutations

S_3 : $(\pm 1, \pm 1, 0, 0)$ with permutations

S_4 : $(\pm 1, 0, 0, 0)$ with permutations

and S_5 : $(0, 0, 0, 0)$.

I would take a combination of these sets. For instance, if I were allowed 24 points, I would use S_1 and S_4, and with 40 points I would use S_1 and S_3 and so on [cf. De Baun, 1959]. Obviously my views have been influenced by both Box's work and by Kiefer's work.

It is, however, also obvious that a realistic procedure should take account of sequential plans. Consider, for example, the investigation of the dependence of a yield variable y on a control variable x in (-1, 1). Suppose that the information on y for each chosen x is available as soon as the experimental run has been made. A rational procedure is not to use Chebychef spacing or Legendre spacing, but to take an observation at x=-1 and at x=+1. One would then take one at x=0, and try to connect three points by a quadratic, or seek a reasonable transformation (non-linear) of the x scale so that the 3 observations fell on a line. One would then probably take additional observations in the middle of the gaps of the best picture one has obtained up to the time of planning new observations. One would, of course, have prefaced the whole matter by obtaining a rough idea of experimental error. It is very difficult to see how the concepts of decision theory and testing of hypotheses can be brought to bear on such a process.

It is clear that practical optimum designing depends on more ingredients than have so far been incorporated in the theory. What one should do depends crucially on:

(a) what use will be made of incomplete information?

(b) what is the rate of feed-back of experimental information?

(c) will the experimenter be able to do additional experiments to fill in gaps in information?

(d) how valuable is information to the experimenter in relation to time? [What is the present value of future information? This will of course depend on what the future information is.]

(e) what is the cost of experimentation? The simple idea of a fixed cost per observation appears to be relevant at best only in some technological studies.

The difficulties of constructing a theory which incorporates these aspects appear to be very great, but should not dissuade us.

FINAL NOTE. It is unavoidable that I cannot describe the results of every paper in the field. The reference list gives only papers referred to and much good work is not discussed. A notable example is the work of Scheffé (1963) on experimentation on a simplex.

REFERENCES

Addelman, S. (1961). Irregular fractions of 2^n factorial experiments. Technometrics 3: 479-496.

__________ (1962). Orthogonal main effect plans for asymmetrical factorial experiments. Technometrics 4: 21-47.

__________ and O. Kempthorne. (1961). Some main effect plans and orthogonal arrays of strength two. Ann. Math. Stat. 32: 1167-1176.

Anscombe, F. J. (1961). Examination of residuals. Proc. fourth Berkeley Symposium on Mathematical Statistics and Probability 1,: 1-36.

__________ and J. W. Tukey. (1963). The examination and analysis of residuals. Technometrics, 5: 141-160.

Blum, J. (1954). Multidimensional stochastic approximation methods. Ann. Math. Stat. 25: 737-744.

Bose, R. C. and N. R. Draper. (1959). Second order rotatable designs in three dimensions. Ann. Math. Stat. 30: 1097-1112.

Box, G. E. P. (1952). Multifactor designs of first order. Biometrika, 39: 49-57.

__________ and D. W. Behnken. (1960a). Simplex sum designs: A class of second order rotatable designs derivable from those of first order. Ann. Math. Stat. 31: 838-864.

__________ and __________ (1960b). Some new 3 level designs for the study of quantitative variables. Technometrics 2: 455-476.

__________ and N. R. Draper. (1959). A basis for the selection of a response surface design. J. A. S. A. 54: 622-654.

__________ and J. S. Hunter. (1957). Multifactor experiments for exploring response surfaces. Ann. Math. Stat. 28: 195-241.

__________ and __________ (1961a). The 2^{k-p} fractional factorial designs. Technometrics 3: 311-352.

__________ and __________ (1961b). The 2^{k-p} fractional factorial designs. Technometrics 3: 449-458.

__________ and K. B. Wilson. (1951). On the experimental attainment of optimum conditions. J. R. S. S., B. 13: 1-45.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. Ann. Math. Stat. 24: 586-602.

Daniel, C. (1959). Use of half-normal plots in interpreting two-level experiments. Technometrics 1: 311-341.

DeBaun, R. (1959). Response surface designs for three factors at three levels. Technometrics 1: 1-8.

Dempster, A. P. (1960). Random allocation designs I: On general classes of estimation methods. Ann. Math. Stat. 31: 885-905.

__________ (1961). Random allocation designs II: Approximate theory for simple random allocation. Ann. Math. Stat. 32: 387-405.

Draper, N. R. (1960a). Second order rotatable designs in four or more dimensions. Ann. Math. Stat. 31: 23-33.

____________ (1960b). Third order rotatable designs in three dimensions. Ann. Math. Stat. 31: 865-874.

____________ (1961a). Third order rotatable designs in three dimensions. Ann. Math. Stat. 32: 910-913.

____________ (1962). Third order rotatable designs in three factors analysis. Technometrics 4: 219-234.

Dvoretsky, A. (1956). On stochastic approximation. Proc. third Berkeley Symposium on Mathematical Statistics and Probability. 439-455.

Ehrenfeld, S. (1955). On the efficiency of experimental designs. Ann. Math. Stat. 26: 247-255.

____________ and S. Zacks. (1961). Randomization and factorial experiments. Ann. Math. Stat. 32: 270-297.

____________ and ________ (1963). Optimal strategies in factorial experiments. Ann. Math. Stat. 34: 780-791.

Elfving, G. (1952). Optimum allocation in linear regression theory. Ann. Math. Stat. 23: 255-262.

Finney, D. J. (1945). The fractional replication of factorial experiments. Ann. Eug. 12: 291-301.

Fletcher, R. and M. J. D. Powell. (1963). A rapidly convergent descent method for minimization. The Computer Journal 6: 163-

Gardner, D. A., A. H. E. Grandage and R. J. Haden. (1959). Third order rotatable designs for exploring response surfaces. Ann. Math. Stat. 30: 1082-1096.

Guest, P. G. (1958). The spacing of observations in polynomial regression. Ann. Math. Stat. 29: 294-298.

Hoel, P. G. (1958). Efficiency problems in polynomial estimation. Ann. Math. Stat. 29: 1134-1146.

Hooke, R. and T. A. Jeeves. (1962). Direct search solution for finding stationary values of a function of several variables. J. Assoc. Comp. Mach., 8: 212-229.

Kesten, H. (1958). Accelerated stochastic approximation. Ann. Math. Stat. 29: 41-59.

Kiefer, J. (1958). On the nonrandomized optimality and randomized non-optimality of symmetrical designs. Ann. Math. Stat. 29: 675-699.

________ (1959). Optimum experimental designs. J. Roy. Stat. Soc. B, 21: 272-319.

________ (1961a). Optimum designs in regression problems II. Ann. Math. Stat. 32: 298-325.

________ (1961b). Optimum experimental designs V with applications to rotatable and systematic designs. Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability 1 : 381-405.

________ (1962). Two more criteric equivalent to D-optimality of designs. Ann. Math. Stat. 33: 792-796.

________ and J. Wolfowitz. (1952). Stochastic estimation of the maximum of a regression function. Ann. Math. Stat. 23: 462-466.

________ and ________. (1959). Optimum design in regression problems. Ann. Math. Stat. 30: 271-294.

________ and ________. (1960). The equivalence of two extremum problems. Canad. J. of Math. 12: 363-366.

Kurkjian, B. and M. Zelen. (1962). A calculus for factorial arrangements. Ann. Math. Stat. 33: 600-619.

Rao, C. R. (1947). Factorial arrangements derivable from combinatorial arrangements of arrays. J.R.S.S., B, 9: 128-139.

Robbins, H. and S. Monro. (1951). A stochastic approximation method. Ann. Math. Stat. 22: 400-407.

Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. Ann. Math. Stat. 29: 373-405.

Satterthwaite, F. E. (1959). Random balance experimentation. Technometrics 1: 111-138.

Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. J. Roy. Stat. Soc. B, 25, No. 2: 235-263.

Shah, B. V., R. J. Buehler and O. Kempthorne. (1964). Some algorithms for minimizing a function of several variables. J. SIAM 12: 74-92.

Tukey, John W. (1959). Discussion of papers by Messrs. Satterthwaite and Budne. Technometrics 1: 166-174.

__________ (1962). The future of data analysis. Ann. Math. Stat. 33: 1-67.

Wald, A. (1943). On the efficient design of statistical investigations. Ann. Math. Stat. 14: 134-140.

Whitwell, John C. and Graham K. Morbey. (1961). Reduced designs of resolution five. Technometrics 3: 459-478.

Wilde, D. J. (1964). Optimum seeking methods. Prentice Hall Inc.

Wilk, M. B. and R. Gnanadesikan. (1963). Graphical analysis of multiresponse experimental data using ordered distances. Technometrics 4: 1-20.

__________ and __________. (1964). Graphical methods for internal comparisons in multiresponse experiments. Ann. Math. Stat. 35: 613-631.

Zacks, S. (1963). On a complete class of linear unbiassed estimators for randomized factorial experiments. Ann. Math. Stat. 34: 751-768.

APPLICATIONS OF DIMENSIONAL ANALYSIS TO MULTIPLE REGRESSION ANALYSIS

David R. Howes
U. S. Army Strategy and Tactics Analysis Group
Bethesda, Maryland

INTRODUCTION. The theory of dimensions which I will discuss, is concerned with the relations that may be found between quantities occuring in nature as a result of the operations which must be performed in order to measure them. Dimensions are things like inches, pounds, minutes, or volts, or rather, the characteriestics which standard measurement units such as inches, pounds, minutes, or volts characterize; namely, length, mass, time, or electrodynamic potential. Physicists and engineers have been making an analysis of these dimensions, as a phase of every problem for many years. The point I want to make today is that a dimensional analysis of a problem should be even more important to a statistician, since such an analysis can reduce both the size of an experiment and the work required to analyze it. As it is not hard to show, a dimensional analysis could, in a given problem, reduce the sample size by more than half. In fact, in the present stage of development of the design of experiments, dimensional analysis offers greater hope for reducing the cost of experiments than any further refinements in construction of blocks, replicates, and so forth. In addition to its promise toward reducing the cost of an experiment, dimensional analysis has another virtue almost equally important. That is, a dimensional analysis carefully conducted can yield a great deal of information, which would otherwise be unobtainable, about the type of model which should be adopted in planning and analyzing an experiment.

Although the basic ideas in dimensional analysis have been in use among physicists and engineers for over a century, they are apparently almost unknown among statisticians; at least there is no reference to the subject in the index of the Journal of the American Statistical Association or any other statistical publication or textbook that I am acquainted with.

However, the theory of dimensions has profound implications in the study of statistical problems. The theory, originated by Joseph Fourier [1] is based upon the observation that; to quote Fourier:

> "Every undetermined magnitude or constant has one dimension proper to itself, and the terms of one and the same equation could not be compared if they had not the same exponent of dimension."

Thus, if a group of variables are connected in a linear equation involving coefficients to be determined by a multiple regression those coefficients must represent quantities whose dimensions are such as to give the same overall dimension to every term in the equation. Similarly also, for equations of higher degree.

Therefore, when linear or polynomial expressions are selected as models for the design or analysis of an experiment, it should be required that any coefficients postulated in these expressions have a dimensionality which bears a reasonable interpretation in context. However, one might justly criticize a model in which one of the coefficients were required to assure the dimension of cubic tons per square degree dollar (and I have seen such an example). If we apply the theory in a more detailed way we can arrive even more exactly at the type of model which should be appropriate, and obtain information concerning those interactions which are to be expected and which can be ruled out.

An example will serve to illustrate what dimensional analysis can provide the statistician. In Duncan, 2, one finds an experiment in which cotton yarn specimens are tested for yarn strength, yarn length, fiber tensile strength, and fineness. Slide No. 1.

X_1: Yarn Strength, Pounds

X_2: Fiber Length, Inches

X_3: Fiber Tensile Strength, Pounds per square inch

X_4: Fiber Fineness, Micrograms per inch.

This problem is discussed and analyzed as one involving one dependent, and three independent variables. However, as a result of dimensional analysis, one is able to postulate:

$$f\left(\frac{X_1 X_3}{X_4^2}\right) = \frac{X_2 X_3}{X_4}$$

where an univariate relationship exists between the quantities on the right and left. An analysis of the data is shown in figure 1. Using the method of least squares, and the data on page 674, one obtains the regression line:

$$X_1 X_3/X_4^2 = .05872 \ (X_2 X_3/X_4) - 3.90$$

with a coefficient of correlation of $r = .955$. Applying this formula to another set of data from the same source given on page 699, and comparing predicted with actual values of X_1, one has a sum of residuals of 107, and a standard error of 9.86. Comparable results using the multiple regression equation given on page 693 are 114 for the sum of residuals and 8.22 for the standard error.

The value of the dimensionless equation is appreciated by considering that it contains only two fitted constants as against four for the multiple regression equation and yet predicts approximately as well. Moreover, the calculations were vastly simplified. Finally, by keeping the number of fitted constants to a minimum, one avoids the danger in complex predictive hyper surfaces that wild contortions may occur in regions which do not happen to be represented in the data, yet which are superficially interpolative. This simplifies and improves the situation from every point of view. In general, the insights provided by dimensional analysis are valuable, and the method is easy.

THEORY OF DIMENSIONS. As is shown in Murray [3], any primary dimension which is effectively present in an experiment or process can be used to reduce the number of variables by one. This fact is explained as follows: External standards of measurement, such as an international metric unit are not necessary to describe a process. Any quantity within the process itself could serve as a standard of measurement for other variables measured in the same dimension. In any formulae, tables or charts describing a process measured in this way, the symbol of the variable taken as the mensurator would not occur, since, being the standard, its value would always be unity. However, an outside observer could convert these same formulae, tables or charts for use with external measurement units, by supplying the symbol of the mensurator as a denominator under the symbol of each variable to measured.

The ratio of a simple or compound variable to its mensurator is referred to as a dimensionless term. Since we reduce the number of variables by one for each primary dimension, m variables in n dimensions can be represented in the form of m-n dimensionless terms provided an adequate system of mensuration can be found.

Each variable may be said to have a vector of dimension

$$P = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_m \end{bmatrix}$$

where each U_i represents the exponent taken by the i th dimension in the dimensionality of the variable as a whole. Thus, if i_1 = mass, i_2 = length and i_3 = time, the dimensional vectors of speed (meters^1 min.$^{-1}$) and pressure (KG^1 meter^{-2}) are:

$$\text{Speed} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \qquad \text{Pressure} = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}.$$

The dimensional vectors of all variables that can be relevant to a problem forms a set which has the property that if a vector P belongs to the set so does CP where C is selected arbitrarily, and if P_1 and P_2 belong to the set, so does $P_1 + P_2$, the vector of the product of the variables. These properties define a linear vector space which is a closed set.

If we can find n variables with linearly independent vectors in this space, these variables are said to span the vector space. The vector of any variable can be duplicated from the n independent vectors by scalar multiplications and vector additions. These independent vectors are a basis for the vector space and a mensurator for any variable can be constructed by combining the variables having these vectors. Any n vectors can be tested for independence by forming the determinant which has these vectors as columns. If it is not zero, they are independent.

Provided then, that a basis of n independent vectors exists, all m-n variables can be measured by mensurators constructed from the n variables having those vectors. Thus, the process can be represented

(1) $$f(\Pi_1, \Pi_2, \ldots, \Pi_{m-n}) = 1$$

where each term is composed of the ratio of a variable to its mensurator.

The theorem above is referred to as the Buckingham Theorem after Buckingham [4]. For practical methods of constructing sets of terms see Langhaar [5], or Murphy [6].

The completely general functional expression (1) is as far as the theory of dimensionality can take us. The explicit function must be determined by experimentation and statistical analysis, or from subject matter theory, or both. When $m - n = 1$, the problem is solved by dimensional analysis alone, and when $m - n = 2$, simple statistical techniques will usually suffice.

MIXED DIMENSIONAL AND DIMENSIONLESS EXPRESSIONS. Previous texts have considered only completely dimensionless representations and have ignored the possibility of a partially dimensionless formulation. Under these circumstances no guidance was provided for the analysis of problems in which the vectors of the variables given are insufficient to span the vector space. This occurs when a complete specification of the forces acting in a process cannot be made. Such incomplete dimensional specifications do not necessarily negate the advantages of dimensional analysis. Some of the variables may still be reduced to a common mensurator, thus, permitting some reduction in the number of variables. For example, consider a chemical experiment with the following variables:

X_1 Amount of Yield, Mols

X_2 Amount of Reactant, Mols

X_3 Amount of Acid, Mols

X_4 Temperature, Degrees, C

X_5 Length of Reaction, Minutes.

Obviously, no mensurator can be found for X_4 or X_5. Therefore, a completely dimensionless expression is impossible - unknown forces have been omitted from the specification. However, X_3 can serve as a mensurator for X_1 and X_2 permitting the formulation

$$X_1 = f \left(X_2, X_4, X_5 \right)$$

where the unit of length is the length of X_3, or

$$\frac{X_1}{X_3} = f \left(\frac{X_2}{X_3}, X_4, X_5 \right)$$

in any units.

Therefore, an incomplete dimensional specification reduces our ability to condense the number of variables. If the variables are all incomensurable we can make no condensation. If, however, some of the variables are commensurable, we can reduce their number to the extent that commensurability exists.

A CHEMICAL WARFARE EXAMPLE. Thus far, we have described a theory which offers a clear-cut reduction in the number of variables required in an experiment. Its implications are so plain that only skepticism concerning its validity would be grounds for ignoring the theory and benefits to be derived from Dimensional Analysis.

In order to dispel skepticism concerning the theory, I have applied Dimensional Analysis to a number of problems in various fields from which data was available; problems in Chemical Engineering, Agricultural Economics, and Quality Control. In every case, the Dimensional Analysis accomplished a successful reduction in the number of variables with a predictive value equal or superior to any previous analysis made using the raw variables.

One application was in the field of assessment of the coverage capability of toxic chemical ammunition against military targets. I am gratified by the results obtained so far, since for many years I was active in this field

and am aware of the high potential savings that would result from any simplification in the problem; especially any model which would eliminate or reduce requirements for testing ammunition over wide ranges of meteorological conditions.

I am aware that much theory has been evolved which purports to describe behavior of toxic clouds in the atmosphere, but also am aware that the mathematical complications of these theories are such that actual models for purpose of prediction rest on approximations whose accuracy is uncertain, and which do not, in my experience, match up with test data obtained in the field, Dimensional Analysis cuts across this theory and leads to an empirical model which accounts for meteorological factors more satisfactorily than existing models.

To illustrate this analysis, Figure No. 1 shows the variables generally agreed to be pertinent to the problem under the assumption of isotropic diffusion. You will note that n, the Sutton turbulence parameter enters into the problem not as a variable, but as the exponent of dimension in which the diffusivity is expressed.

The temperature is omitted from this list since there is no completely agreed manner for considering it and it does not fit into the dimensional picture here. Sutton's theory ignores it and it is customary to consider it as a component of source strength; varying the effective source strength.

Figure 1 shows a set of three dimensionless Π terms which according to our theory should be able to replace the six variables shown on the previous slide. A study of these terms shows that the data from one experiment in the field could be used to predict the results of other experiments under different meteorological conditions. Also, it implies that the results of all conceivable experiments could be represented by a single surface in three dimensional space, or as a family of curves in two dimensional space.

Figure 2 shows the results of two field trials plotted in the space of the dimensionless variables shown previously. The two trials were conducted with the same type of shell, and at approximately the same temperature. However, the wind and stability conditions were considerably different, and therefore, the coverage figures obtained were also considerably different. In the 202 trial on the left the wind speed was 1 meter/sec as correspond with 3.23 meters/sec for the trial 203 the right. Stability was moderate inversion for the trial on the left and moderate lapse for that on the right.

The Sutton parameters n, and C were calculated form from the wind-height profiles given in the test reports using the Barad-Hilst equations.

As the chart shows, the two trials were sufficiently different to prevent any overlap between the two families of curves. However, the critical point is that the two sets of observations are recognizably members of the same family and that the curve - 8.6, which occurs in both sets of data matches up very well: in fact, a line projected through the two points obtained in trial #202 passes exactly through 4 of the 6 points shown for trial #203. This is highly encouraging since it was only in meteorological conditions that the trials were different, implying that the analysis given did, in fact, satisfactorily account for the changes in the area dosage curve, and did so for every time interval.

We infer from this example that additional tests could be analyzed to fill in the blank spots on our chart and an empirical equation fitted to this data with ease, since only three variables are involved, and the curves obtained are approximately colinear.

DIMENSIONAL ANALYSIS AND MULTIPLE REGRESION ANALYSIS. Dimensional Analysis is a great help in solving the difficulties encountered in multiple regression analysis. It has several advantages:

a. The number of variables, and therefore the extent of the calculations required, is reduced.

b. The freedom with which alternative representations of the data can be formed facilitates the discovery of collinear representations which simplify the analysis.

c. The predictive equation partakes of a structural validity not entirely dependent on statistical estimation.

The value of the dimensional approach may be appreciated in relation to the Bean Ezekiel graphical method of multiple curvilinear regression analysis, [7]. In that procedure, no explicit mathematical form need be ascribed to the relationship among the variables but by an iterative graphical process an increasingly accurate approximation to the curves involved is obtained, and the result is a set of charts which can be used directly for predictive purposes, or, if desired, converted to tables,

nomographs or slide rules. A scatter plot of residuals is also obtained, for an estimate of error. The principal drawback of the method was the frequent inability of the analyst to isolate recognizable "draft lines" at the outset due to non-collinearity of response. The freedom of dimensional representation should largely overcome this difficulty and increase the scope of the method.

CONCLUSION. The foregoing exposition has shown that the application of dimension theory to statistical problems can result in valuable insight and savings in experimental design and analysis and should, therefore, become part of the equipment of statisticians generall. Objections to the theory have at times been advanced, usually on the basis of special examples wherein functional invariance under change of units prevails without dimensional homogeniety (see Bridgman, [8]). However, in its favor, the results obtained by Dimensional Analysis are obtained also from the theory of partial differential equations as applied to physical problems (see Langhaar, Chapter 10); the theory has successfully supported the researches of Maxwell, Rayleigh, Helmholtz, and others, and neither the literature nor the experience of the present writer offers an instance wherein the supposed relationships have been found absent in fact.

It is also unclear to what extent the standard statistical designs, tests, and techniques customarily applied to dimensional variables can be applied to dimensionless variables. Thus, it is recognized that many developments in error analysis and the theory of sampling will be required to exploit Dimensional Analysis to its fullest (a recent paper by Halperin and Mantel, [9] would appear to be of value in this connection). An obvious case requiring attention is that of setting confidence limits on a dependent variable which is a constituent of one or more terms, although setting limits for the term themselves would be straightforward.

It is hoped that being made aware of the advantages of Dimensional Analysis, statisticians will bend it to their needs with the necessary developments.

REFERENCES

1. Fourier, J. The Analytical Theory of Heat Dover Press, 1955 Ch. II, Sec. 9.

2. Duncan, A. J. Quality Control and Industrial Statistics Richard Irwin Co. 1959, p. 674.

3. Murray, F. J. Mathematical Machines Vol. II, Columbia University Press 1961, p. 229.

4. Buckingham, E. On Physically Similar Systems Phys. Rev. Vol. IV, No. 4, 1914, p. 345.

5. Langhaar, H. L. Dimensional Analysis and the Theory of Models Wiley & Sons 1951, Ch. 4.

6. Murphy, G., Similitude in Engineering Ronald Press, New York 1950. p. 23.

7. Ezekiel, M. and Fox, K. Methods of Correlation and Regression Analysis Wiley and Sons, 1951. Ch. 16.

8. Bridgman, P. W. Dimensional Analysis Yale University Press, 1922, pp. 41-43.

9. Halperin, M. and Mantel, N. Interval Estimation of Non-Linear Parametric Functions, JASA, Vol. 58, No. 303, 1964, pp. 611-627.

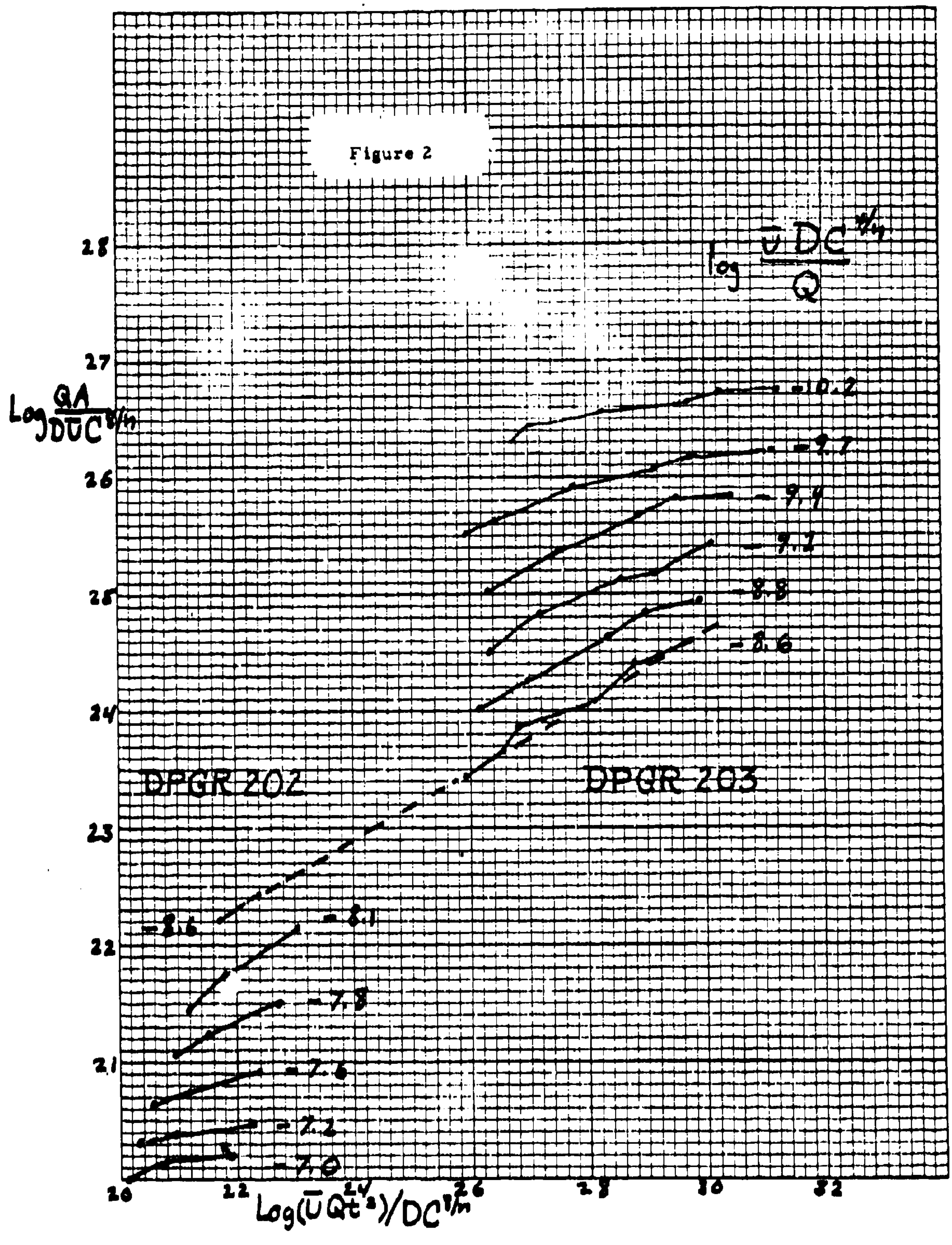
CHEMICAL WARFARE PROBLEM

Variable	Dimension
AREA	$MTRS^2$
DOSAGE	$MG\text{-}MIN/MTR^3$
SOURCE STRENGTH	MG
TIME	MINUTES
WIND SPEED	MTRS/MIN
DIFFUSIVITY	$(MTRS)^{n/2}$

$$\frac{QA}{DUC^{8/n}}, \quad \frac{\bar{U}DC^{4/n}}{Q}, \quad \frac{\bar{U}Qt^2}{DC^{8/n}}$$

Figure 1

Figure 2


log $\frac{\bar{U}DC^{3/4}}{Q}$
-10.2
-9.7
-9.4
-9.2
-8.8
-8.6
DPGR 202
DPGR 203
-8.6
-8.1
-7.8
-7.6
-7.2
-7.0
28
27
26
25
24
23
22
21
20
22
24
26
28
30
32
Log $\frac{QA}{D\bar{U}C^{3/n}}$
Log$(\bar{U}Qt^2)/DC^{1/n}$

THE USE OF REGRESSION ANALYSIS FOR CORRECTING FOR MATRIX EFFECTS IN THE X-RAY FLUORESCENCE ANALYSIS OF PYROTECHNIC COMPOSITIONS

Raymond H. Myers
Virginia Polytechnic Institute: Blacksburg, Virginia

and

Bernard J. Alley
U. S. Army Missile Command: Redstone Arsenal, Alabama

I. INTRODUCTION. X-Ray fluorescence methods are widely used in industry for the analysis of a variety of materials. The non-destructive nature and exceptional speed of these methods are largely responsible for their widespread use and increasing acceptance. The direct analyses of many materials, for example can be accomplished 20 to 50 times faster than by conventional chemical procedures. This allows sufficient time after an analysis to permit the correction or rejection of a substandard batch of material before processing is completed.

The actual X-Ray fluorescence method of analysis may be briefly described as follows: the primary beam from an X-Ray tube impinges on the surface of a specially prepared sample. The components in the sample surface are immediately excited and emit their characteristic emission lines in all directions. Qualitative analyses are made by determining the angles at which the characteristic emission lines from the sample occur. Quantitative analyses can in general be performed on a particular component of a mixture, of say K components by positioning the radiation detector at an angle which corresponds to the characteristic emission line for that component and measuring the emission line intensity. The intensity is then related to the component percentage by a suitable calibration procedure.

The intensity of a component's characteristic radiation is not a simple function of the concentration of that component alone in the sample. The intensity depends also on the concentrations of the other components. This is caused by the absorption and enhancement among the components, of the primary and excited radiation. The existence of these interelement or "matrix" effects is one of the more serious problems encountered in X-Ray fluorescence analysis and hence inhibits, to a great extent, the use of this technique as a quantitative analytical tool.

Many non-mathematical methods have been devised to either minimize or correct for these matrix effects. However, they have been found to be either too costly or too time consuming on samples from large scale production of multicomponent mixtures. It is the purpose of this paper to discuss the use of regression analysis in the correction of these interelement effects for the estimation of concentration of individual components in a mixture and to emphasize the application to a particular solid rocket propellant mixture in current use by the U. S. Army at Redstone Arsenal, Huntsville, Alabama.

Effect of Solid Particle Size

A problem which may be encountered when one is analyzing slurry mixtures containing solid constituents is the influence of solid particle sizes on the X-Ray intensities. It might be necessary that any analytical procedure contain some type of correction for this effect, unless of course the individual particle sizes always remain constant throughout production. Part I of this paper gives the analytical technique for the situation in which the particle sizes were experimentally held constant. Part II extends the analytical procedure to the case of variable particle size.

II. ESTIMATION OF CONCENTRATION (Particle Size Constant). Samples of a five component solid propellant mixture were prepared and analyzed for four of the components. (The actual ingredients are classified and hence we shall denote them in the text as components 1, 2, 3, and 4 respectively). These samples were taken from the twelve different batches in a narrow concentration range in which the product is usually manufactured. The particle sizes of the solids in the slurry mixture were held essentially constant. The number of seconds for a fixed count intensity measurements were recorded in rapid succession for each component. The same was done for a synthetic standard sample. The response variable used was $R = t_s/t_c$, where t_s is the number of seconds for the standard and t_c the number of seconds for the component in question. This is standard procedure used in this type of X-Ray work. The purpose of the standard and the subsequent use of the ratio of the standard reading to the unknown reading is to correct for electronic and mechanical changes in the spectrograph. The data is found in Table I.

Consider the model;

$$R_{ij}=B_{i,0}+B_{i,1}X_{1j}+B_{i,2}X_{2j}+B_{i,3}X_{3j}+B_{i,4}X_{4j}+\epsilon_{ij}, \quad (i=1, 2, 3, 4) \tag{1}$$

where R_i is the intensity ratio for component i, X_1, X_2, X_3, and X_4 are the concentrations of the individual components, the B's are regression coefficients, and ϵ_{ij} is the random error associated with R_{ij}. Note that the concentrations of each component appear in the model despite which of the four ingredients is being detected. Least squares estimates of the regression coefficients were found for the four regression lines. These estimates are shown in Table II along with the error mean squares for the regression lines. The intensity measurements are not in general linearly related to concentration but in the reasonably narrow range of interest shown in Table I, a linear relationship appears to hold quite well.

TABLE I. Intensity Ratio Measurements and Composition of Mixtures

(Compositions in weight percent)

Batch	R_1	R_2	R_3	R_4	X_1	X_2	X_3	X_4
1	1.1240	0.8980	0.8219	0.9906	0.5514	70.18	12.53	15.04
2	0.9285	0.8872	0.9308	0.9944	0.4426	68.84	14.26	14.75
3	1.1214	0.8030	0.7668	1.1221	0.5631	67.51	12.79	17.39
4	1.1635	0.8706	0.9272	0.9832	0.5624	67.52	14.83	15.34
5	0.9415	0.8064	0.9026	1.1127	0.4505	66.10	14.52	17.03
6	0.9039	0.8314	0.7596	1.0994	0.4425	68.86	12.30	16.72
7	1.0712	0.8404	0.8662	1.0836	0.5290	67.34	13.95	16.35
8	0.9561	0.8731	0.8206	1.0290	0.4702	69.00	13.07	15.68
9	1.0186	0.8431	0.8346	1.0591	0.5001	68.07	13.51	16.02
10	1.0744	0.8124	0.7432	1.0967	0.5379	68.52	12.24	16.64
11	0.9005	0.8320	0.8606	1.0798	0.4321	67.26	13.93	16.34
12	0.9318	0.8913	0.8126	0.9880	0.4498	69.96	12.49	14.99

TABLE II. Regression Coefficients and Error Root Mean Squares

Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4
s_e=0.00768	s_e=0.00776	s_e=0.01130	s_e=0.01265
$b_{1,0}$=0.15411	$b_{2,0}$=-1.4370	$b_{3,0}$=-1.51670	$b_{4,0}$=0.60788
$b_{1,1}$=1.8573	$b_{2,1}$=0.01832	$b_{3,1}$=-0.07426	$b_{4,1}$=-0.13257
$b_{1,2}$=-0.00074	$b_{2,2}$=0.03020	$b_{3,2}$=0.02008	$b_{4,2}$=-0.00442
$b_{1,3}$=0.00919	$b_{2,3}$=0.02561	$b_{3,3}$=0.08024	$b_{4,3}$=-0.00641
$b_{1,4}$=-0.00832	$b_{2,4}$=-0.00790	$b_{3,4}$=-0.00328	$b_{4,4}$=0.05605

We can use the equations in (1) to develop a set of working expressions for estimating the concentrations, i.e.,

$$\underline{R}=\underline{b}+B\underline{\hat{X}} \tag{2}$$

where $\underline{R}$ represents the vector of intensity ratios and $\underline{b}$ the vector of intercept terms. The b_{ik} element of B is the coefficient of X_k in the $\underline{ith}$ regression line. $\underline{\hat{X}}$ is the vector of unknown concentrations that one seeks to estimate in practice. Inverting (2), we have:

$$\underline{\hat{X}}=B^{-1}(\underline{R}-\underline{b}). \tag{3}$$

Here we have a case of the use of a set of simultaneous multiple linear regression lines in reverse, i.e., inverting the regression lines to estimate the X's i.e., the concentrations. Williams [3] gives a discussion of the general problem. It might be noted that the concentrations were used as the independent or concomitant variable since the error in the X's is very small as compared to that for the X-Ray intensity ratios.

Equation (3) represents a working set of equations for estimating the concentration from samples from running production. The four equations given by the matrix expression in (3) are as follows:

$\hat{X}_1 = -0.14381 + 0.54061\ R_1 + 0.07935\ R_2 - 0.08034\ R_3 + 0.08670\ R_4$

$\hat{X}_2 = 38.2619 - 0.5767\ R_1 + 42.5690\ R_2 - 13.1116 R_3 + 5.1478\ R_4$

$\hat{X}_3 = 8.9016 + 0.6984\ R_1 - 10.4829\ R_2 + 15.6926\ R_3 - 0.4547\ R_4$

$\hat{X}_4 = -7.1523 + 1.3131\ R_1 + 2.3448\ R_2 + 0.5705\ R_3 + 18.4010\ R_4$

The residual errors of estimation, calculated from the original data, are shown in Table III.

TABLE III. Residual Errors of Estimation of Concentration

Batch	$X_1 - \hat{X}_1$	$X_2 - \hat{X}_2$	$X_3 - \hat{X}_3$	$X_4 - \hat{X}_4$
1	-0.0035	0.02	-0.19	-0.09
2	0.0026	0.43	-0.14	-0.23
3	0.0013	-0.01	0	0.10
4	-0.0026	-0.04	0.14	0.30
5	-0.0026	0.16	-0.24	0.07
6	-0.0026	0.03	0.06	0.07
7	0.0027	-0.30	0.01	-0.31
8	0.0046	-0.42	0.24	0.13
9	0.0016	0	0.12	-0.11
10	0.0011	0.39	-0.06	-0.13
11	-0.0014	-0.17	0.11	0
12	-0.0012	-0.14	-0.02	0.19

CONFIDENCE INTERVAL ESTIMATES ON THE CONCENTRATIONS. Box and Hunter [1] discuss the problem of joint confidence interval estimates on the solution of a set of simultaneous equations when the coefficients are subject to error. Their work was actually a part of a more specific problem of finding a confidence region for a stationary point in response surface analysis. However, the procedure also applies to our problem of attaching confidence limits to concentrations. Suppose that in general we have K simultaneous equations of the type;

$$(4) \qquad \sum_{j=0}^{K} b_{ij} X_j = 0 \qquad (i=1, 2, \ldots. K)$$

where the b_{ij} are subject to error (for our case $X_0=1$). Consider the quantities,

$$\sum_{j=0}^{K} b_{ij}\, \xi_j = \delta_i \ (i=1, 2, \ldots\ldots K),$$

where the ξ are the values of the X's that would satisfy (4) if the actual regression coefficients were used in place of the b_{ij}. If we consider a vector of the δ's, say $\underline{\delta}$ as having a multivariate normal distribution with mean vector $\underline{0}$ and variance-covariance matrix $E(\underline{\delta}\,\underline{\delta}')=V$, then the expression $\underline{\delta}'V^{-1}\underline{\delta}$ follows a X^2 distribution [2] with K degrees of freedom. For our case, the ith element of $\underline{\delta}$ can be written as $R_i-\hat{R}_i$, where $\hat{R}_i$ is the estimate of the intensity ratio in the ith regression line. For estimates of the elements of V, we can write

$$\mathrm{Var}\,(R_i-\hat{R}_i)=s_{ii}\,[1+\frac{1}{n}+\sum_h\sum_l C_{hl}\,\xi_h\xi_l]$$

$$=s_{ii}.H.$$

$$\text{and } \hat{\mathrm{Cov}}\,(R_i-\hat{R}_i,\ R_k-\hat{R}_k)=s_{ik}[1+\frac{1}{n}+\sum_h\sum_l C_{hl}\,\xi_h\xi_l\,]$$

$$=s_{ik}.H.$$

where:

s_{ii}= sample estimate of the variance of R_i for particular values of $\xi_1, \xi_2, \xi_3, \xi_4$.

s_{ik}= sample estimate of the covariance between R_i and R_K.

C_{hl}=(hl) element of the inverse of the matrix of corrected sums of squares and products of the X's for the calibration sample.

If we replace the elements in V by their corresponding estimates and divide by the appropriate degrees of freedom we arrive at the ratio

$$\frac{n-8}{4} \sum_i \sum_k \frac{\delta_i \delta_k w^{ik}}{H}$$

which is distributed as F with 4 and n-8 degrees of freedom, where w^{ik} is the (ik) element of the inverse of the matrix W, the matrix of residual sums of squares and products of the R's. We can write

$$\begin{aligned} \delta_i &= R_i - \hat{R}_i \\ &= \sum_j b_{ij} \hat{X}_j - \sum_j b_{ij} \xi_j, \end{aligned} \tag{5}$$

where the $\hat{X}_j$ are the estimates of the concentration obtained from equation (3). If we replace δ_i by the expression in (5), we have

$$\begin{aligned} F_{4,n-8} &= \left(\frac{n-8}{4}\right) \frac{\sum_i \sum_j \sum_k \sum_l (\hat{X}_j - \xi_j)(\hat{X}_l - \xi_l) b_{ij} b_{kl} w^{ik}}{H} \\ &= \left(\frac{n-8}{4}\right) \frac{\sum_j \sum_l (\hat{X}_j - \xi_j)(\hat{X}_l - \xi_l) q_{jl}}{H} \end{aligned} \tag{6}$$

where q_{jl} is the (jl) element of the matrix;

$$Q = B' W^{-1} B.$$

Here b_{ij} is the (ij) element of the matrix B.

Equation (6) represents simultaneous joint confidence interval estimates of the actual concentrations ξ_1, ξ_2, ξ_3, and ξ_4. Thus if we are given values of the estimates $\hat{X}_1$, $\hat{X}_2$, $\hat{X}_3$, and $\hat{X}_4$, we can substitute particular values of the concentrations ξ_1, ξ_2, ξ_3, and ξ_4 into equation (6) and if the resulting

expression is less than $F_{\alpha, 4, n-8}$ (upper tail), then those values of the ξ's fall inside the $100(1-\alpha)\%$ confidence band.

The elements of the W^{-1} and Q matrices are:

$$W^{-1} = \begin{bmatrix} 7214.8162 & 2554.8459 & -3201.5790 & 3439.8046 \\ & 4014.0983 & -1714.2325 & 2122.4663 \\ & & 2679.7650 & -1867.4456 \\ & & & 2825.4942 \end{bmatrix}$$

$$Q = \begin{bmatrix} 24274.424 & -15.343 & -274.4115 & 219.582 \\ & 2.4899 & 2.8058 & 0.77389 \\ & & 10.8662 & -3.6781 \\ & & & 5.3264 \end{bmatrix}$$

III. VARIABLE PARTICLE SIZE. An experiment was conducted in a manner similar to that described in II except that the particle size was allowed to vary. Components 2 and 4 are the only ones for which the particle size is an important factor in its effect on the intensity ratio measurement. The point should be made here that it is assumed that the particle sizes are known in a practical situation, i. e., for a sample of the propellant from running production one can determine, from the physical source of components 2 and 4, at least the mean particle size. The degree of difficulty here would depend upon the precision with which these two components are manufactured. No attempt was made here to consider such problems as particle size distribution. Likewise no attempt was made to consider the degree to which the particle sizes of components 2 and 4 are altered by the mixing process itself.

A 1/8 fraction of a 2^6 factorial design was used with four replications at each point and in the center of the design. The factors are the concentrations X_1, X_2, X_3, X_4, and particle sizes W_2 and W_4. Table IV gives the design matrix and the defining contrasts.

TABLE IV. Design Data and Defining Contrasts

Batch	Treatment Combination	X_1	X_2	X_3	X_4	W_2	W_4
1	abef	1	1	-1	-1	1	1
2	cdef	-1	-1	1	1	1	1
3	(1)	-1	-1	-1	-1	-1	-1
4	ace	1	-1	1	-1	1	-1
5	bde	-1	1	-1	1	1	-1
6	abcd	1	1	1	1	-1	-1
7	adf	1	-1	-1	1	-1	1
8	bcf	-1	1	1	-1	-1	1
9	midpoint	0	0	0	0	0	0

Defining contrasts: I, ADE, BCE, ACF, BDF, ABCD, ABEF, CDEF. (Particle Size Units are per cent fine fraction on total ingredient basis)

A set of multiple regression equations of the type

$$R_{ij} = \sum_{k=0}^{4} \left[B_{ik} X_{kj} \right] + B_{i5} W_{2j} + B_{i6} W_{4j} + \epsilon_{ij} \quad (i=1,2,3,4) \tag{8}$$

were fit to the design data, where as before $X_0=1$. Table V shows the estimates of the coefficients of the regression line in (8). (8) can be written as

$$\underline{R} = B_1 \hat{\underline{X}} + B_2 \underline{W}.$$

We can then "correct" the intensity ratio vector for particle size and solve for the vector $\hat{\underline{X}}$;

$$\hat{\underline{X}} = B_1^{-1}(\underline{R} - B_2 \underline{W}) \tag{9}$$

This results in the following set of equations

$$\hat{X}_1 = 11.998725R_1 + 598.526R_2 + 82.076R_3 + 395.848R_4 - 988.676 - 13.5897W_2 - 2.2816W_4.$$

$$\hat{X}_2 = 8.84359R_1 + 3207.192R_2 + 439.287R_3 + 2109.9897R_4 - 5226.75124 - 74.7551W_2 - 11.4927W_4.$$

$$\hat{X}_3 = 1.653744R_1 + 867.2777R_2 + 137.368R_3 + 578.007R_4 - 1437.2059 - 19.37799W_2 - 3.02207W_4$$

$$\hat{X}_4 = 3.0437R_1 + 1258.126R_2 + 175.7089R_3 + 847.6307R_4 - 2073.6127 - 28.46016W_2 - 5.32504W_4.$$

The equation in (8) could also be used to estimate particle size when either the particle size cannot be determined or one feels that the mixing process has caused sufficient "grinding" that there has been a change from the particle sizes of the pure components. Of course this would require a chemical analysis of two of the components of the mixture, which of course, is time consuming.

TABLE V. Estimates of Regression Coefficients and Error Root Means Squares for Equation (8)

Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4
$s_e = 0.02005$	$s_e = 0.01199$	$s_e = 0.00830$	$s_e = 0.02298$
$b_{10} = -4.8413$	$b_{20} = 2.82710$	$b_{30} = -8.4503$	$b_{40} = -8.19590$
$b_{11} = 1.9320$	$b_{21} = -0.03948$	$b_{31} = 0.11398$	$b_{41} = 0.08438$
$b_{12} = 0.05104$	$b_{22} = -0.01436$	$b_{32} = 0.09337$	$b_{42} = 0.08200$
$b_{13} = 0.06237$	$b_{23} = -0.02355$	$b_{33} = 0.15847$	$b_{43} = 0.08462$

TABLE V
(cont'd.)

Ingredient 1	Ingredient 2	Ingredient 3	Ingredient 4
b_{14}=0.05010	b_{24}=-0.05424	b_{34}=0.07888	b_{44}=0.14812
b_{15}=-0.00582	b_{25}=0.01072	b_{35}=-0.00682	b_{45}=-0.00815
b_{16}=0.00024	b_{26}=-0.00218	b_{36}=-0.00245	b_{46}=0.00417

Table VI shows the residual errors in estimation of the concentration using equation (9).

TABLE VI. Residual Errors in Estimation of the Concentration Using Equation (9) (Units in wt. %)

Batch	Ingredient 1 $\hat{X}_1-X_1$	Ingredient 2 $\hat{X}_2-X_2$	Ingredient 3 $(\hat{X}_3-X_3)$	Ingredient 4 $(\hat{X}_4-X_4)$
1	-.006332	.2216	-.1006	-.0196
2	-.008813	.3308	-.1491	-.0373
3	-.013621	.5426	-.2567	-.0684
4	.003560	-.0692	.0417	-.0485
5	.008362	-.2675	.1365	-.0293
6	-.001603	.0227	-.0061	.0097
7	.004032	-.0966	.0539	-.0513
8	.007943	-.2583	.1234	-.0280
9	.007812	-.8720	.3427	.3712

SUMMARY. A set of equations is given for estimating the component concentration in a certain solid propellant mixture in terms of the X-Ray intensity readings of each component. The method used involves inverting a set of simultaneous multiple linear regression equations. The concentration of each ingredient appears in each equation in order to correct for "matrix" conditions which do effect the X-Ray intensities. The significance tests on

individual components indicate that these interelement conditions do, in fact, exist for the mixture in question. Joint confidence regions were developed for the concentrations.

Since it was suspected that the particle size of pure components 2 and 4 also effect the X-ray intensity, a linear model involving particle size was fit to the data from a 1/8 fraction of a 2^6 factorial design. This did indicate that particle size was in fact a necessary consideration and resulted in a set of equations for estimating the concentration of each component in terms of an intensity reading which is adjusted for particle size.

ACKNOWLEDGEMENTS. Thanks is given to Mr. Robert Davis who helped with the calculations and Mr. Robert Tankersley of the U. S. Army Missile Command Computation Center, Redstone Arsenal who performed the necessary regression analyses.

BIBLIOGRAPHY

(1) Box, G. E. P. and Hunter, J. S. (1954). A Confidence Region For the Solution of a Set of Simultaneous Equations With an Application to Experimental Design. Biometrika, 41; 190-199.

(2) Graybill, Franklin A. (1961). "An Introduction to Linear Statistical Models." McGraw-Hill Book Company, Inc.

(3) Williams, E. J. (1959) "Regression Analysis". John Wiley and Sons Inc.

SAMPLING FOR DESTRUCTIVE OR EXPENSIVE TESTING

Joseph Mandelson
Quality Evaluation Division, Quality Assurance Directorate
Edgewood Arsenal, Maryland

INTRODUCTION. In recent years the engineer has been impressed with the fact that the principles of sampling are essentially statistical in character because the effect of sampling can only be appraised in terms of operation of the laws of chance. Consistent with this revelation, the engineer by and large has been content to retire from the field of sampling and abdicate his responsibilities in this area to the statisticians. A few hardy souls, confirmed do-it-yourselfers, took it upon themselves to invade the statistical field and learned to acquit themselves creditably in the area of sampling. They even branched out into other aspects of statistics germane to engineering. However, the influx of engineers into the statistical preserve was not sufficiently large to be able to handle the relatively heavy volume of activity required. Then, too, a number of working tools were prepared by statisticians, presumably for use by quality engineers and inspection personnel, to cover a multitude of sampling problems as these occur in quality assurance. Some of these tools are quite complicated; for their complete comprehension they demand more in the way of statistical knowledge on the part of the would-be user than the authors are prepared to admit. As a consequence there is a degree of obscurity in the field. The engineer is urged to consult the statistician whenever his state of confusion or the importance of the matter in hand appears to warrant. However, the engineer should long algo have risen in wrathful protest against statistical tools supposedly prepared for his use but which he finds slippery and elusive to the point of unintellgibility.

Actually, is it so important that comprehension of the mathematical derivation of statistical methods be made an essential prerequisite to their efficient use? Would not an explanation of the basic factors, in non-mathematical terms followed by a detailed by-the-numbers procedure to use in the given context suffice? At any rate, I propose to try this approach.

SAMPLING RISKS. The layman has long regarded the field of sampling with healthy suspicion; he has felt in his bones that sampling is a risky business at best. The well-publicized failures of public opinion polls in predicting the results of crucial election weaken his confidence in statistical methods. His instinct in regard to risks is, of course, entirely

correct as his everyday experience with matters governed by the law of chance illustrates. Allusion may be made to games of change, insurance and the like. Much can be learned from pertinent analogies. Let us consider examples from games of chance such as bridge, etc.

A well shuffled card deck is analogous to a lot of material from which a sample is taken, with one important difference: the exact composition of the deck is known, that of the lot is not. Each hand in bridge is a sample of 13 from a lot of 52. A hand of exactly average strength would contain one card of each of the 13 values and a 4-3-3-3 distribution in suits. Our experience tells us that such a hand is almost never observed. Instead we find that some hands are stronger and, by the same token, others are weaker than the average. This should teach us that a sample is very rarely truly indicative of the composition of the lot. Instead, we find that the sample sometimes appears to be better, sometimes worse than the average, if by average we mean a sample whose composition is exactly proportionate to that of the lot. Further, we find that small variations from the average-strength hand are quite frequently encountered, large variations are relatively rare.

In real life, the composition of the lot is almost never known, the purpose of the sample is to permit us to make inferences and decisions regarding the acceptability of the lot sampled. Since we recognize that the sample rarely reveals precisely what the true quality is, it must be accepted that some of the decisions at which we arrive, based on results obtained in testing the sample, may be in error. There are two types of such error.

<u>PRODUCER'S RISK</u>. The Type I error, so called, is the decision to reject a lot which is really acceptable. This occurs when the sample, through chance variation, indicates a larger proportion of defectives than that which is really present in the lot. It is the equivalent of the bridge hand which contains almost no strength. These hands occur occasionally, with predictable frequency. In the same way lots of acceptable quality will produce a sample of given size which, with predictable frequency, will indicate the lot to be unacceptable. It should be noted that, while the frequency of such occurrences can be predicted (say once in twenty samples) the actual event (which one, if any, of the twenty) cannot be foreseen; it occurs at random intervals. In any case, the rejection of an acceptable lot occurs with a certain probability equivalent numerically to this predicted frequency of the Type I error. Since a rejected lot will require 100% inspection of the lot,

rework or scrapping of the material, it is plain to see that the risk of this unfortunate occurrence is one which will cause the producer some economic loss. For this reason this is called the Producer's Risk.

CONSUMER'S RISK. On the other side of the coin we have the Type II or beta error which occurs when we decide to accept a lot which is really unacceptable. This occurs when the sample, by chance, yields results which happen to conform to the requirements which decide the acceptability of material offered him. This situation is analagous to the bridge hand which is abnormally strong. The comments already made with respect to the Type I error are also applicable to the Type II error, viz., the frequency of such occurence can, within reason, be predicted if certain information, normally not available, is at hand or can be assumed. The effects of the Type II error are quite different, of course, since the material now becomes the property of the user and the excessively high proportion of defectives it possesses will undoubtedly cause him to sustain certain kinds of loss. The Type II error gives rise to the Consumer's Risk.

EFFECT OF SAMPLE SIZE ON RISK. Both types of error and the associated risks may be reduced by using larger samples. It can be shown that the amount of information concerning the quality of the lot, available from the sample, varies as the square root of the numerical size of the sample. Consequently, if one wishes to double the information in the sample he must multiply his sample size by four. Clearly, this can soon become an expensive business and leads to diminishing returns.

It must ever be kept in mind that the risks we have considered have substantial significance, economic and otherwise. Both risks lead to various types of loss, many (but not all) of which can be measured in monetary terms and all of which must be assumed either by the supplier or the consumer. Whether these costs will weigh more heavily on the former or the latter is determined by the quality of the lot, the sampling plan and the level of quality specified. The risks and, therefore, their cost can be reduced by increasing the sample size but this, in turn, raises the cost of sampling and test which is customarily borne by the consumer. We are reminded that raising the sample size to effect an arithmetic increase in information will necessitate a geometric increase in the costs associated with the sample size.

TOTAL COST OF SAMPLING. If one is realistic he will recognize that the total cost of sampling includes not only the cost of taking and testing the sample but also the losses occasioned by the operation of the risks already discussed. It may appear strange, perhaps unbelievable, that there should be any who will not accept the fact that there are risk losses to evaluate and will not agree to include these in the reckoning. But these doubting Thomases are like their predecessor - unless they see little green bills passing over a counter from one hand into another they cannot agree that a cost or loss has been sustained. It is particularly unfortunate when such short-sighted persons get into a position where they are able to influence the sampling plan to be used. When, in consequence, losses are sustained from defectives regarding which complaints are received from users, and from lots unnecessarily screened or reworked, such people eloquently display newly washed hands as tokens of their freedom from sin and learnedly discuss the poor inspection job turned out by that overly-large and over-paid staff of inspectors. Now, say these management experts, if we really want to save money, here is some fat which can be advantageously trimmed. It will never occur to them that insistance on minimum sample sizes reduces a relatively small cost but incurs much larger risks which require the piper to be paid in large and repeated installments.

The true total cost of sampling is determined by several parameters, chief among which are the sample size, the specified quality level, and the consumer's and producer's risks. There are other parameters involved in the final result such as the cost of making a test, the true quality of the lot, the cost of reworking an item declared defective, etc. For our purpose, it is desirable to search out the interrelationships among the four parameters first mentioned.

Clearly, the larger the sample, the more costly the test. At the same time, the risks and their attendant costs are reduced by large samples. This situation leads naturally to the supposition that there may be some point at which the size and cost of the sample are so happily related to the costs of the corresponding risks that the over-all cost is a minimum. The size of sample which, within the stated conditions, brings about such a desirable result may, with propriety, be designated the optimum sample size. The existence of such an optimal solution can easily be demonstrated arithmetically (2). However, there are some matters which we should clarify before venturing further. These include the meaning of and ways to handle the cost of the risks.

COSTING THE PRODUCER'S RISK. The producer's or alpha risk has already been described as the risk that the sample may indicate the lot to be unacceptable when it is, in fact, quite acceptable. If the test is non-destructive or the cost of making the test is not prohibitively high, it is economically possible to test or examine each item in all rejected lots. In this way the original erroneous decision will be corrected at a price - the cost of such test or inspection is the cost of rejecting the lot and, under these circumstances, the price paid for the Type I error is relatively low. But if the test is quite expensive, particularly if it damages or destroys the item tested, it is not feasible to test each item in the lot. Hence a rejection, whether right or wrong, is practically an order to scrap the lot or rework it. In this case, the cost of the producer's risk is painfully evident especially when one recalls that the producer's risk causes rejection of acceptable lots which, due to a sampling quirk, give the false impression of being rejectable. In any case, the cost of rejecting a lot is easy to calculate and it is given in the following symbolic form: (The meaning of the symbols is provided in the Glossary appended hereto.)

$$C_R = (N - n)(C_U - V_S)(P_P)$$

It should be obvious that C_R, the cost of rejection, can be computed to the last penny; very few approximations are necessary.

COSTING THE CONSUMER'S RISK. It is otherwise with the task of calculating the cost of the consumer's risk in dollars and cents. We will recall that the consumer's risk is the chance he takes that the sample may represent an unacceptable lot as acceptable material. This causes him to pay for and take possession of merchandise which contains an undesirably high proportion of defective material. There the difficulty begins; to assess the cost of accepting a defective lot one must solve the problem of fixing the cost of a single defective item and follow this by discovering the actual percentage of defectives in the lot. If the latter information were at hand, it would have been unnecessary to test the lot for acceptability in the first instance and, had the test revealed the true percent defective in the lot, it would never have been accepted. This difficulty pales to insignificance compared with the problem of determining the cost of an item found to be defective when it is used. This is particularly true of exotic items such as space rockets and military material where failure in use may have strong adverse effect on national prestige and/or security, may cause casualties or even lead to tactical defeat in situations of various degrees

of significance. Almost always the loss due to the defective unit depends upon the circumstances surrounding the malfunction. These are unpredictable. Thus, a premature shell burst may cause no casualties or damage in certain situations or it may result in several deaths and a ruined gun. Chance, completely unforeseenable, will determine the loss in each case. Again, how can we compare the cost of a dud hand grenade on the practice field with the loss sustained when a grenade, thrown into an enemy machine gun emplacement, is a dud and the brave soldier who had to expose himself to the gun to make the throw, is cut down? Someone else will have to make that throw and who can tell how many casualties will be sustained to silence the gun which would have been destroyed had the grenade functioned in the first place? The additional casualties are part of the loss associated with the dud. How can anyone predict the course of such events? If one wishes to dramatize this problem he may say that his objective is to put a price on human blood and look into his crystal ball to determine, on the average, how much will be poured out on each defective item.

We must not take the attitude that the cost of the beta risk can never be ascertained. If the item involved is a component and the defect is one that will be caught in attempting to assemble it in the end item then the nuisance loss of this type of defect can be determined. In that case, the method described in Reference (1) can be used for determining sample size while minimizing the total cost of both risks and of sampling.

As we shall see later, the cost of the two risks strongly influence the sample size determined to the optimum in the sense of reducing the total cost to a minimum. If the cost assessed therefore is very high, the optimum sample size calculated to reduce the total cost to a minimum will be unrealistically high as will the minimum total cost computed in these circumstances. In a democracy such as ours, great value is placed on human life. It is commonly regarded as priceless and any attempt to set a monetary value on blood or on life itself is considered a particularly obnoxious form of sacrilege. Yet if such matters are to enter in to the calculation of optimum sample size in a specific case, a monetary value must be set. The engineer seems to be impaled on the horns of an insoluble dilemma.

HOW TO HANDLE THE CONSUMER'S RISK. Yet a solution is possible. The price of blood or life must simply be equated to zero. In other words, it must be eliminated from consideration in monetary terms as suggested

in Reference (3). Such a step makes the problem soluble. In this case, the casualty-producing defective can better and more appropriately be handled by prescribing a suitable quality level for acceptance. To adopt this course is equivalent to a decision to eliminate the casualty-producing defective in its role of a sample size determinant and to direct its influence into another path, so that it will act to determine the pertinent quality level instead.

LOT TOLERANCE. One way to handle the problem of determining the optimum sample size for destructive tests, without assessing any cost for the consumer's risk (this is the same as ignoring it or setting it equal to zero) is provided in Reference (2). There the required quality level is set at a figure appropriate to the protection desired as an LTFD (= Lot Tolerance Fraction Defective; see Glossary) which is a level of quality so poor that the engineer would take to his sick bed at the thought of having to accept consistently material of LTFD quality though, once in a long while, to prevent shutting down the line or for some other noble purpose, he might be willing to accept such a lot. By setting the Consumer's risk at some low figure (e.g. 0.10 or 0.05) the engineer insures that only one lot of LTFD quality out of 10 or 20 submitted will be accepted, the others being rejected. Obviously no producer can stand the economic pressure of wholesale rejection, so the quality he must produce to stay in business will have to be a good deal better than the LTFD, which is what our engineer wants. Having decided on a proper LTFD the paper goes on to show how the optimum sampling plan is computed which will yield the desired protection against material of LTFD quality.

Reference (1), on the other hand, is a much more sophisticated approach. However, as has already been noted, it can be applied only where the cost of the beta risk can be computed with reasonable correctness, at least to the extent of knowing in what ball park the doubleheader will be played. Our concern, however, is with the area within which the cost of the beta risk cannot be approximated. It is interesting that the solution herein delineated can be used equally well whether one can or cannot estimate the beta risk cost because in either case the cost can be ignored, if desired, and the acceptance or surveillance quality level may be set at a figure which will keep the outgoing lot percent defective at some desired limit with given probability given some information as to distribution of lot quality. That is, we set the level to take a calculated risk. Then we figure the sampling plan that will insure that outgoing material accepted thereby will conform to that level within the specified risk.

Now we shall consider how this purpose may be accomplished by the engineer without the need to become a statistician, amateur or professional. To do this, we propose to outline the procedure "by the numbers" and ask the engineer to accept as an article of faith that the procedure is, in fact, valid and will do the things and afford the protection attributed to it. It is not our purpose to provide mathematical theory or proofs here and demand that you grasp them before we will permit you to touch the procedure. Rather, we want to present a method which you can grasp in hands grimy from contact with your work and responsibilities and from a knowledge of your problem and needs, proceed to calculate a sampling plan tailored to do what you want it to do.

COMPUTING ACTUAL COSTS. If we consider the case of single sampling (see Glossary) wherein we fix the consumer's risk (i.e. by establishing some desired lot tolerance fraction defective with a 10% chance of acceptance - the consumer's risk), the total cost of the inspection is expressed by the equation

$$T = n\,(C_U + C_T) + (N - n)\,(P_P)\,(C_U - V_S)$$

Since this equation is basic to understanding what we are about to do, it is well to explain it without going to the Glossary. T is the total cost of testing including the Producer's Risk the cost of which is the expression to the right of the central plus sign. To the left of that sign is the cost of testing: n, the sample size, times the sum of the cost of one unit (which the test will destroy) and the cost of testing it. Thus, if the sample size is 35 and we shall destroy an item costing \$3 and spend \$2 to do it, then the test alone will cost 35 x (3 + 2) = \$175. Now, as for the Producer's Risk, the rest of the lot, N - n, is subject to the probability (P_P) that it will be rejected even though the lot is really acceptable. The symbol P_P is the Producer's Risk; it is computed as $1 - L\bar{p}$ by subtracting from unity the chance, $L\bar{p}$, that a lot of process average quality ($\bar{p}$), presumably better than LTFD, will be accepted. If unity represents all possibilities and $L\bar{p}$ is calculated as a decimal fraction, say, 0.95 then $1 - L\bar{p}$ is the chance of rejection; in this case 1 - 0.95 = 0.05. Now (N - n) (P_P) gives the number, on the average, which we will lose from the lot by the action of the Producer's Risk. We may not lose this lot but when we do lose a lot and its N - n is prorated over all the lots we do not lose, each lot will

lost about $(N - n)(P_p)$. It remains only to cost this loss. This is done by multiplying $(N - n)(P_p)$ by the cost of one item less its salvage value, if any, $C_U - V_S$. If an item costs \$10 and can be reworked for \$3, then $C_U - V_S =$ \$3 so that $(C_U - V_S)$ may also be called the cost of reworking the item. When the appropriate values are filled in, the total cost T of using any proposed sampling plan against material of the quality being produced ($\bar{p}$) may be calculated. A bit laborious but, as you can see, not too difficult.

The calculation, from scratch, of an optimum sample size would require quite a bit of work. First, as indicated in (2), one would have to determine a succession of different sample sizes and an associated allowable number of defects (c) for each. Each plan must be designed to furnish the same protection (same Consumer's Risk) against material of lot tolerance (LTFD) quality. Then, the total cost of each plan would be computed, using the above equation. It would require facility in using a table of probabilities. While this would not be difficult to learn, such a table is, after all, a statistician's reference. Happily, Ellner and Savage (4) have developed short-cut methods for calculating optimum single and double (see Glossary) sampling plans utilizing graphical methods and graphs developed by Dodge and Romig (5). These graphs are reproduced and appended hereto with the kind permission of the originators and publishers and, in any case, can be consulted in (5).

THE WORK OF DODGE AND ROMIG. It is generally acknowledge that Dodge and Romig are the fathers of statistical sampling as used in quality assurance work. It is astonishing to see how sophisticated their thinking was, even in its earliest published form in the Bell Telephone Technical Journal. Their methods are intensely practical but that should not surprise anyone since they were engineers faced with the eminently practical problem of sampling. While their rejected lots would be inspected 100%, they recognized that the cost of such 100% inspection is an economic loss. Their sampling plans were calculated to minimize the over-all cost of the inspection operation including the 100% inspections caused by the Producer's Risk. Therefore the idea of optimizing sample size for minimum cost originated with Dodge and Romig. The use of the same principle for destructive or expensive testing where 100% inspection of rejected lots was patently impracticable was urged by (2) and (4), substituting $C_U - V_S$ for the Dodge and Romig's 100% inspection of rejected lots. With this great similarity in basic ideas, it is not too surprising that we can use Dodge

and Romig's graphical methods to avoid a good deal of computational work which might be not only laborious but confusing to the non-statistician. To avoid the latter, we propose to develop single and double sampling plans using the Dodge-Romig graphs and to proceed step by step explaining only as required to facilitate achievement of the final objective - the sampling plan.

CONTROLLING THE PROCESS. In their eagerness to insure receipt of high quality material, engineers can easily fall into the trap of specifying acceptance criteria so high as to increase production and inspection costs beyond reason and hamper production of a smooth flow of acceptable material. For the dubious advantage of an exceedingly low outgoing proportion of defective material, the consumer pays through the nose. There are other ways to do this without incurring prohibitive costs and strangling production. Perhaps the most effective way is to engineer production and establish effective quality controls at the right points on the production line so that production of the most critical or significant types of defects will be almost impossible. Another way, not as effective and more costly, but easier and more convenient for the purchaser is to establish an LTFD at such a level that, to avoid a costly high proportion of rejections, the producer will have to maintain an average quality output well above the LTFD.

ESTABLISHING THE LTFD. In establishing the LTFD we shall assume a Consumer's Risk of 10% or 0.10 for two reasons. First, ever since Dodge and Romig first calculated their tables this has been the risk conventionally accepted for the LTFD. Second, their graphs are based on an 0.10 risk. The engineer should set his LTFD at some fraction defective such that, even if a lot of LTFD quality were accepted on rare occasion, it would cause no insurmountable problem in the field. Since sampling plans developed by our method with reject lots of LTFD quality nine times out of ten, if the contractor would regularly produce material of this quality he would surely face economic disaster. If the supplier's Producer's Risk is to be at a tolerable level he must produce material by a process which is statistically controlled to give a process average ($\bar{p}$) proportion defective very roughly 1/3 or 1/4 of the LTFD. Thus, if the LTFD is 0.08, the supplier should produce a $\bar{p}$ of about 0.02 or 0.03 to avoid excessive loss due to the Producer's Risk. If the supplier's $\bar{p}$ is much lower than the LTFD the optimum sample size will be relatively low.

The engineer should choose an LTFD that will give him what he needs at an acceptable price. From the facts already indicated, he must have a reasonable expectation that the supplier will be able to produce a controlled $\bar{p}$ which is 1/3 LTFD. If he cannot, his prices will have to be raised to cover the excessive rejections he is sure to experience. The engineer must avoid demanding material of prohibitively high quality solely for the purpose of bolstering his reputation for designing items which work all the time. He must remember that, if the supplier is trying to make material at a controlled $\bar{p}$ = 1/3 LTFD, very rarely will the process make a lot of LTFD quality and, even if it does, the chance of its being accepted is only one in ten, so the engineer can rest assured that, for practical purposes, almost all accepted lots will be much better than LTFD quality. With this in mind, he can afford to be fairly generous in setting the LTFD.

Perhaps as good a way as any is to assume some realistic $\bar{p}$ which the engineer feels a qualified supplier can maintain under statistical control when producing the item in question. Then the engineer multiplies $\bar{p}$ by 4 and 3 and asks whether a product of quality $4\bar{p}$ or $3\bar{p}$ can, on rare occasion, be accepted without causing excessive trouble to the user. Using this as a criterion he sets his LTFD at $4\bar{p}$ if possible, at $3\bar{p}$ otherwise. The engineer should realize that, if the supplier maintains control over his quality a lot of LTFD quality will almost never be produced, much less accepted. The supplier should recognize that if a sampling plan is computed on an LTFD basis he would be well advised to get his process under statistical control at a $\bar{p}$ no greater than 1/3 LTFD and keep it there. If, for some reason, the LTFD must be set at some figure noticeably less than $3\bar{p}$, the engineer should expect higher prices, uncertain deliveries or repeated requests for waivers or changes in contract requirements. The supplier can anticipate occasional, even frequent rejections and organize with this possibility in mind. The above procedure is only a useful rule-of-thumb. By making a number of trial calculations, the engineer can satisfy himself that when $\bar{p}$ is very small compared with LTFD, the sample size required will be relatively small and rejections will be few. As $\bar{p}$ approaches the LTFD, sample size will be at a high and rejection will tend to occur in 9 cases out of 10.

DESIGNING THE OPTIMUM SINGLE PLAN (EXAMPLE 1). To illustrate how to design a single sampling plan, we shall use the example furnished in (4). First we shall list by symbols the things we need to know

quantitatively. If any of this information is lacking, it is advisable to use your best guess and make any correction which later information indicates to be suitable.

$N = 5000$	$C_U = \$5$
LTFD $= p_t = 0.07$	$C_T = \$10$
$\bar{p} = 0.02$	$V_S = \$3$

We calculate the qualtities $A = C_U + C_T = \$15$ and $B = C_U - V_S = \$2$. Usually A and B can be determined quite accurately but they are not as important as the ratio $\frac{B}{A}$. Using these figures, we calculate the following:

$$p_t N = 0.07 \times 5000 = 350$$

$$\frac{B}{A}N = \text{approximate equivalent lot size} = \frac{2}{15} \times 5000 = 667$$

$$\frac{\bar{p}}{p_t} = \frac{0.02}{0.07} = 0.29, \text{ and}$$

$$p_t \frac{B}{A}N = (p_t)\,(\text{approximate equivalent lot size}) = 0.02 \times 667 = 46.7.$$

We enter Figure 2 with $p_t \frac{B}{A}N = 46.7$ and $\frac{\bar{p}}{p_t} = 0.29$ and get an acceptance number $c = 4$. Now going to Figure 3, we follow the curve for an acceptance number of 4 and we find it leaves the chart at p_tN of 200. Our p_tN is 350 but since the curves for $c = 0$ to $c = 10$ remain parallel to the horizontal axis past $p_tN = 200$, we read (p_t) (sample size) or $p_tn = 8$. Since $p_t = 0.07$ we find $n = \frac{8}{0.07} = 114$. We substitute 114 in the expression for the exact equivalent lot size, $p_t[\frac{B}{A}N + (1 - \frac{B}{A})n]$ which converts to $0.07[\frac{2}{15} \times 5000 + (1 - \frac{2}{15})114] = 53.6$. We could not calculate the exact equivalent lot size before this because we need to know n, the sample size. That we obtained by first using the approximate equivalent lot size. We re-enter Figure 2

with the new estimate of n_t (equivalent lot size) = 53.6 and $\frac{\bar{p}}{p_t} = 0.29$ and get c = 5. Now we re-enter Figure 3 with c = 5 and $p_t'N = 350$ and read 9.2 (by using dividers and a scale). Since $p_t = 0.07$, $0.07n = 9.2$ whence n = 131. The optimum single sampling plan, then, is n = 131, c = 5. We can check this by recalculating the long expression above and getting 54.5 which when used to enter Figure 2 again with $\frac{\bar{p}}{p_t} = 0.29$, finds c = 5 unchanged. That is all there is to it.

INFLUENCE OF THE PROCESS AVERAGE, $\bar{p}$. To insure that the optimum in sampling economy is maintained, the process average should be recomputed every 5 or 10 lots. If any sizeable change is noted, it would be wise to recompute the sampling plan, which is not an onerous task as you have seen. The question may be put as to what value to use for $\bar{p}$ when calculating the original sampling plan, when no quality history exists for the production line. At such a time, your best guess as to the average quality the line is expected to produce is adequate or you may prefer to estimate $\bar{p}$ conservatively at about $0.3p_t$. It probably will not make too much difference either way since, even if the estimate is off somewhat, it will not be too far away and will be changed as soon as a quality histroy becomes available. As an exercise, one might vary the process average, using some figures much higher and much lower than $\bar{p} = 0.02$ and notice the effect on the sample size which results from the change.

DOUBLE SAMPLING. Some time ago, double sampling and the related multiple sampling were regarded as ways to reduce the over-all cost of sampling since, for sampling plans giving the same protection the total number of sample items needed for single sampling was normally noticeably more than what double sampling demanded which, in turn, was greater than what multiple sampling required. Thus, if the amount of retesting could be kept down, as when quality is either very good or very poor, appreciable savings appear possible. Since the system for calculating optimum single sample plans takes into account changes in sample sizes when $\bar{p}$ changes, it possesses some of the advantages of double and multiple sampling without the disadvantages. Again, many like the idea of getting a second chance with double sampling, several chances with multiple sampling. One does not feel so tied down to the one chance of the single sample. This is, of course, purely psychological for, mathematically, there is a price to pay. Additional costs must be borne in selecting second and other samples that

are used only infrequently. There is the physical burden and inconvenience of handling more sample items and of returning unused samples to the parent lots. Then, too, when retests become more frequent than originally anticipated, heavy work loads are experienced leading to over-work, fatigue and, eventually, to error. These factors have caused double and multiple sampling to lose some of their popularity and led to greater dependence upon and use of single sampling plans. Nevertheless, we shall include a method for computing optimum double sampling plans.

DESIGNING THE OPTIMUM DOUBLE SAMPLING PLAN (EXAMPLE 2). For this example we shall use the figures used in Example 1. To spare you the trouble of looking them up they are listed below:

$N = 5000$	$C_U = \$15$
$LTFD = p_t = 0.07$	$C_T = \$10$
$\bar{p} = 0.02$	$V_S = \$3$

Again we calculate $A = C_U + C_T = \$15$ and $B = C_U - V_S = \$2$. Using these figures we calculate

$$p_t N = 0.07 \times 5000 = 350$$

$$\frac{B_N}{A} = \text{approximate equivalent lot size} = \frac{2}{15} \times 5000 = 667$$

$$\frac{\bar{p}}{p_t} = \frac{.02}{.07} = 0.286 \text{ and}$$

$$p_t\, B_N/A = (p_t)\,(\text{approximate equivalent lot size}) = 0.02 \times 667 = 46.7.$$

To determine the respective c numbers for our double sampling plan we use Fig 2-7 which is analagous to Fig 1-2. We enter Fig 2-7 with $p_t\, B_N/A = 46.7$ for the ordinate or vertical component and $\bar{p}/p_t = 0.286$ for the horizontal component or abscissa. We find $c_1 = 1$ and $c_2 = 7$, almost inside $c_2 = 8$.

Now we use Fig 2-8 and, at $p_t N = 350$, the curve for $c_1 = 1$ gives a reading of 4.5 on the ordinate which represents $p_t n$ or p_t times the first sample size. Since $p_t n_1 = 4.5$ and $p_t = 0.07$, $n_1 = \frac{4.5}{0.07} = 64$. Similarly we look up c_2 for $p_t N = 350$ and we find an ordinate of 12.8 which now represents $p_t (n_1 + n_2)$. Now if $p_t (n_1 + n_2) = 12.8$ and $p_t = 0.07$ then $n_1 + n_2 = \frac{12.8}{0.07} = 183$. Since $n_1 = 64$, $n_2 = 183 - 64 = 119$. As before, this is a first approximation to the sampling plan we want. Substituting in the expression

$$\frac{B_N}{A} + (1 - \frac{B}{A})(n_1 + n_2)$$

we get $667 + (1 - 2/15)(183) = 826$. Back we go to Fig 2-7, using $p_t (826) = 0.07 \times 826 = 57.8$ and we get $c_1 = 1$ and $c_2 = 8$. Again we enter Fig 2-8 with $p_t N = 350$ as the abscissa and for $c_1 = 1$ we get $p_t n_1 = 4.5$ so that $n_1 = 64$ as before. However for $c_2 = 8$, we get $p_t (n_1 + n_2) = 14.0$ whence $n_1 + n_2 = \frac{14.0}{0.07} = 200$, from which $n_2 = 200 - 64 = 136$. The sampling plan then is $c_1 = 1$, $c_2 = 8$, $n_1 = 64$, $n_2 = 136$. If desired, the sample sizes can be rounded to $n_1 = 65$, $n_2 = 135$ without too great a change in the effect of the plan. As you can see, the calculations are a bit more involved for the double sampling plan as compared with the single sampling plan but the principle is the same.

The desire to keep the presentation simple requires omission of several facets which might be useful such as an easy way to calculate the expected total cost of a given sampling plan if β is known. However, if this information is required it can be obtained from other graphs in (5).

In all the previous discussion, it was assumed that the only information available regarding the quality of the lot to be tested was that developed from the sample. In an actual production situation a substantial amount of engineering information is developed during the production

cycle which, properly interpreted, can indicate whether the process is in statistical control and, therefore, may be considered to be producing substantially homogeneous material. If the material is homogeneous from lot to lot then the results of tests generated in previous lots may be considered to have significant bearing on the results expected in the latest lot. Hence when statistical control has been established, the sample size, lot by lot, can be reduced substantially and remain reduced provided no evidence is obtained indicating loss of control.

Basically, if advantage is taken of available engineering knowledge of previous experience with the process sampling, testing, and their attendant costs may be reduced. This notion lends itself readily to statistical ingenuity but the engineer will require the assistance of a statistician to take advantage of the possibilities. A number of ingenious schemes to permit useful employment of existing engineering data can be devised to reduce the sample size and test costs below the "optimum" solution just described.

The author desires to express his appreciation and gratitude to Ellner and Savage for permission to use the results of their research and most particularly to Professor Harold F. Dodge, Dr. Harry G. Romig, and John Wiley and Sons, Inc. for their unselfish generosity in allowing reprinting of their graphs without which this work would have been impossible.

REFERENCES

(1) Barnard E. Smith, "Some Economic Aspects of Quality Control", Applied Mathematics and Statistics Laboratories, Stanford University, Technical Report No. 53, 3 July 1961.

(2) Joseph Mandelson "Estimation of Optimum Sample Size in Destructive Testing by Attributes", Industrial Quality Control, November 1946.

(3) E. G. D. Paterson "Quality Control Engineering in Product Evaluation", Industrial Quality Control, May 1960 wherein the author indicates that cost cannot intelligently be assigned to the beta risk and that this factor can best be governed by "...the employment of acceptance criteria and procedures which will, to the extend practicable, obviate their presence in the accepted product." Paterson was vice-president of Bell Laboratories in charge of quality control.

(4) H. Ellner and I. R. Savage "Sampling for Destructive or Expensive Testing by Attributes" presented at the Second Engineering Statistics Symposium at Army Chemical Center, Md., in April 1956 and at the Army Science Conference, West Point, N. Y., in June 1957.

(5) H. F. Dodge and H. G. Romig, Sampling Inspection Tables, 2nd Ed., John Wiley and Sons, New York, 1959.

(6) Joseph Mandelson "Lotting", Industrial Quality Control, May 1962.

GLOSSARY

C_R = Cost of rejection

N = Lot size

n = Sample size

C_U = Cost of a single unit

V_S = Salvage value of a single unit or its value as rework material

P_P = Producer's risk: probability (expressed as a decimal fraction) that the sample will, on test, represent the lot to be unacceptable when it is, in fact, quite acceptable

C_S = Cost of sample item

C_T = Cost of testing a single unit

$A = C_U + C_T$ = The cost of destroying one item in testing

$B = C_U - V_S$ = The value of one rejected item

c = Acceptance number, the maximum number of defectives that will be permitted in a sample of size n from an acceptable lot. If more than c defectives are observed in the sample of n items the lot will be rejected.

DOUBLE SAMPLING SYMBOLS

n_1 = Size of first sample

n_2 = Size of second sample

$n_1 + n_2$ = Size of combined first and second samples

c_1 = Acceptance number for first sample, n_1. If c_1 or fewer defectives are found in n_1, the lot is accepted straight-away. If the number of defectives found in n_1 is greater than c_1 but equal to or less than c_2, the second sample, n_2, is tested and the number of defectives in n_1 and in n_2 is totalled. If that number is greater than c_2 (the number of defectives permitted in $n_1 + n_2$) the lot is rejected. If c_2 or less defectives are found in $n_1 + n_2$ on retest, the lot is accepted.

DEFINITIONS

Single Sampling - A system of sampling whereby a single sample is drawn from a lot and the acceptability of the lot is determined from the results obtained in testing the sample. No retest is permitted if results are unfavorable.

Process average ($\bar{p}$) - The apparent proportion of percent of defectives manufactured by the production process. It is generally computed by dividing the total number of defectives found in the samples taken from the last few lots tested (5 or 10) by the sum of the sample sizes. This gives $\bar{p}$ as a decimal fraction.

Lot Tolerance Fraction Defective (LTFD or p_t) - Lot quality, expressed as a decimal fraction defective, so poor that we want to permit only a small chance or probability (the Consumer's Risk, say one chance in 10 = 10% = 0.10 probability) that the sampling plan will permit acceptance if such a lot is submitted.

Double Sampling - A system of sampling wherein two samples are taken and one set of acceptance and rejection criteria are furnished for each sample. If the results obtained in testing the first sample meet neither the acceptance nor the rejection criterion for that sample, the second sample is tested (called the retest) and the decision is made using the second set of criteria. A decision is always possible using the second set of criteria after the retest.

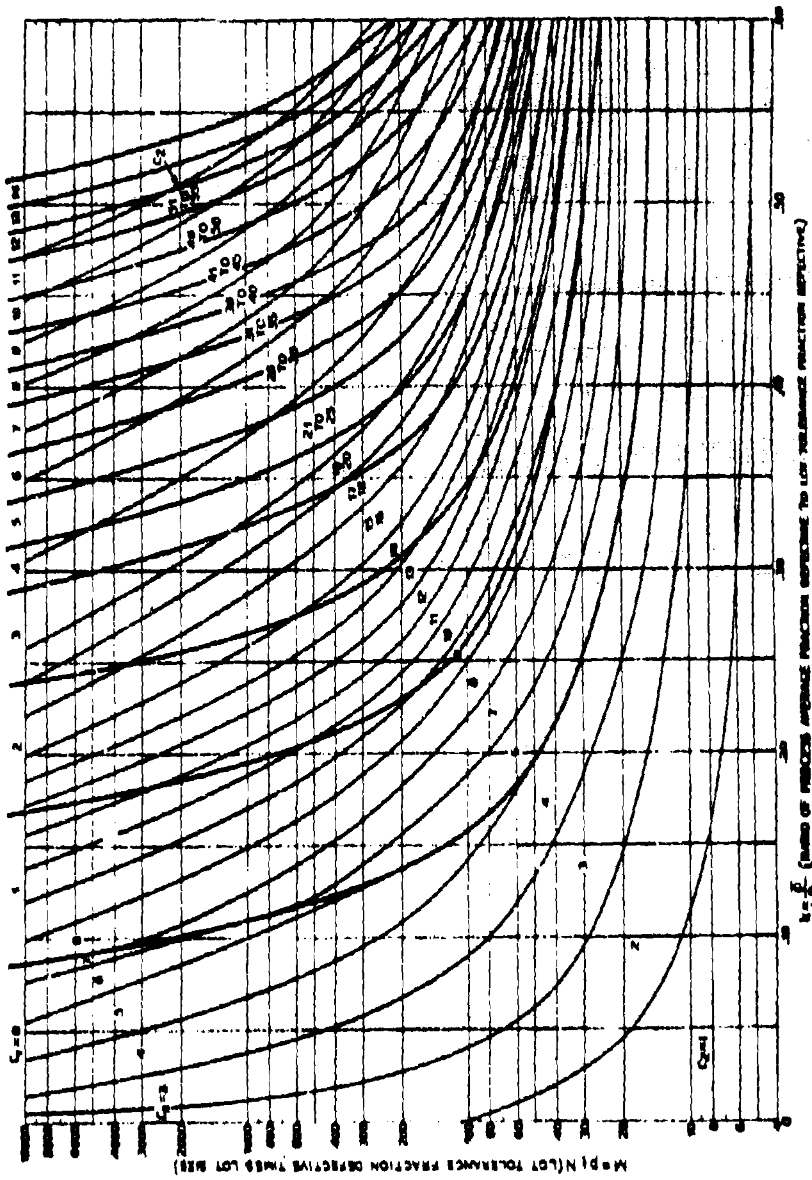


Fig. 2-7 Chart for determining acceptance numbers c_1 and c_2 for lot tolerance protection, Consumer's Risk 0.10

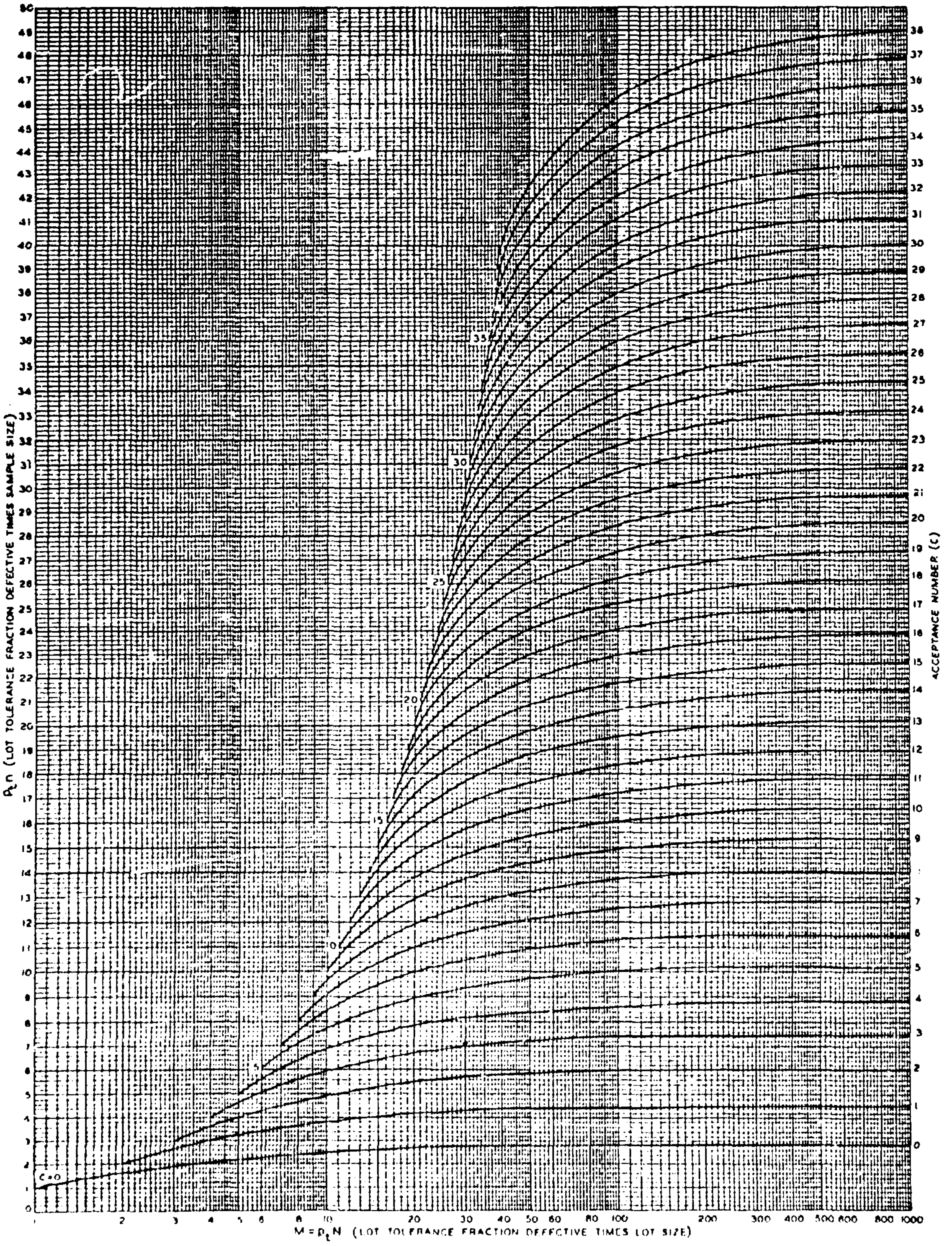


Fig. 2.8 Chart for determining sample sizes n_1 and n_2: lot tolerance protection, Consumer's Risk 0.10

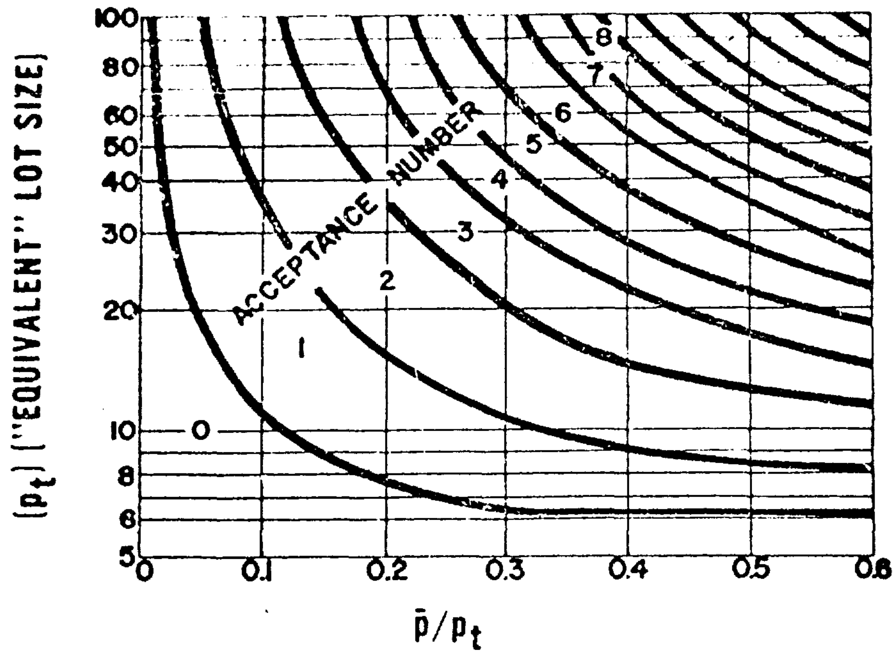


FIG. 2.* CHART FOR FINDING ACCEPTANCE NUMBER OF SINGLE SAMPLING PLAN. (CONSUMER'S RISK, 0.10).

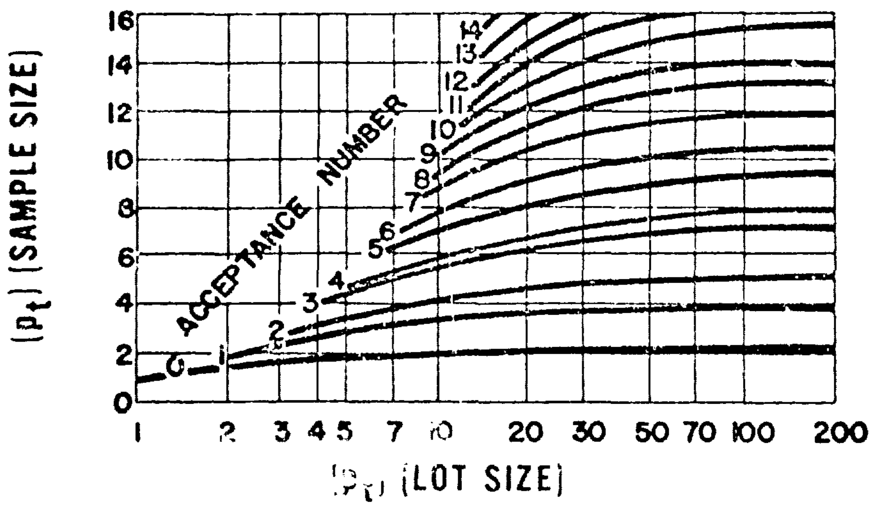


FIG. 3.* CURVES FOR FINDING SIZE OF SINGLE SAMPLING PLAN. (CONSUMER'S RISK, 0.10).

*Reproduced in part, by permission, from "Sampling Inspection Tables" by Dodge & Romig, published by John Wiley & Sons, Inc.

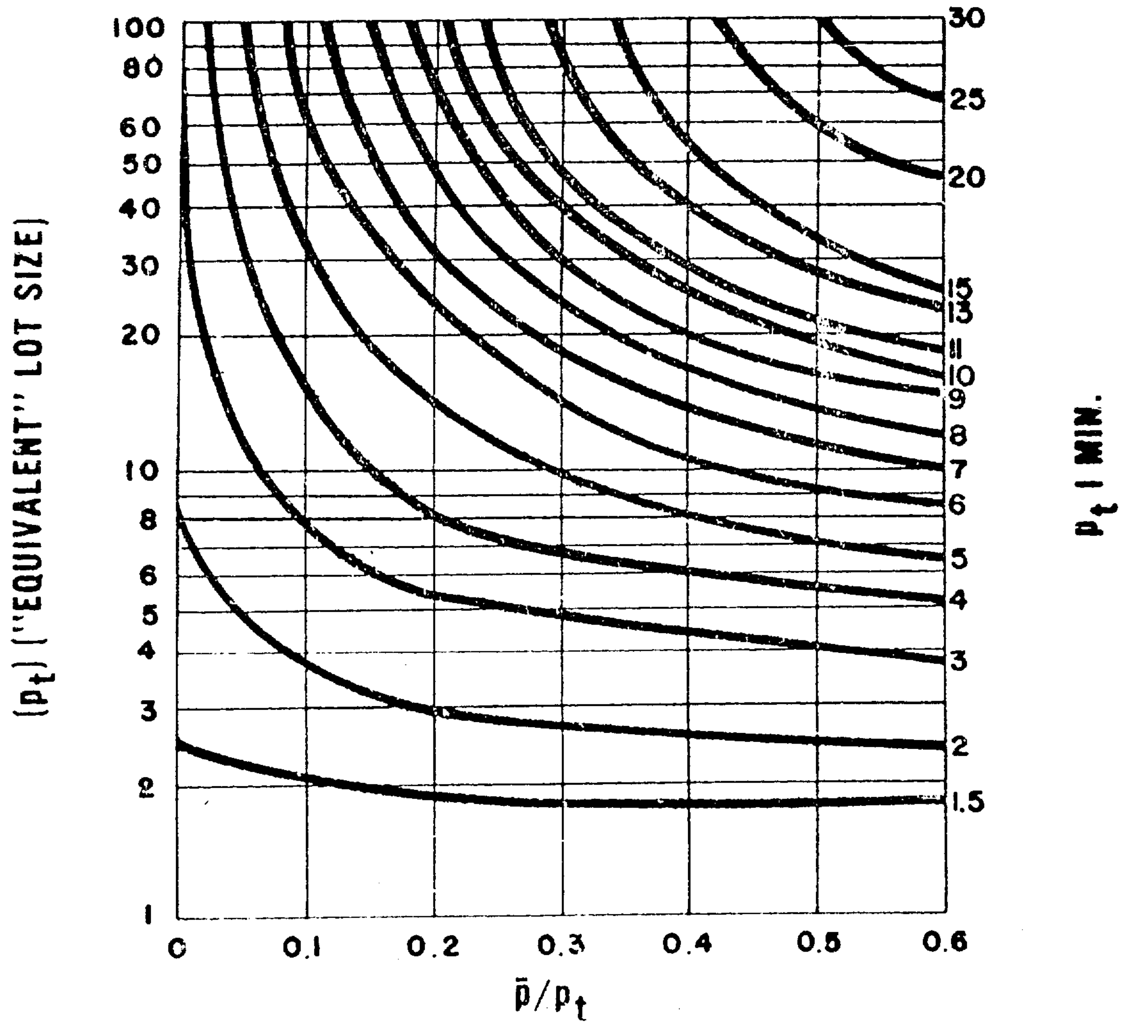


FIG. 4.* CURVES FOR FINDING THE MINIMUM COST OF INSPECTION PER LOT (SINGLE SAMPLING PLAN - CONSUMER'S RISK, 0.10)

*Reproduced in part, by permission, from "Sampling Inspection Tables" by Dodge & Romig, published by John Wiley & Sons, Inc.

PROCEDURES FOR FINDING TOTAL SAMPLE STATISTICS FROM SUBSAMPLE STATISTICS

Paul C. Cox
Reliability and Statistics Division
Army Missile Test and Evaluation Directorate
White Sands Missile Range, New Mexico

ABSTRACT. While procedures for obtaining the variance for a total sample from subsample statistics is fairly well known, there appear to be very few instances in which such procedures are found in print. Therefore, twenty-five formulas are presented which are in one way or another. related to obtaining the mean and variance for a total sample from subsample statistics. In addition, techniques are demonstrated for using these formulas to determine the mean and variance for a sample in which a portion of the observations have been modified, some have been added, or a few have been deleted.

The discussion includes: applications of these formulas; precautions which should be observed; methods for deriving the formulas; and, procedures for their use.

I. INTRODUCTION. This report presents techniques and formulas for determining the mean and variance of a total sample if this sample has been partitioned into a set of non overlapping and mutually exhaustive subsamples; and the mean, variance, and sample size are known for each subsample.

Similarly, techniques are discussed for changing the variance when observations are added to, deleted from, or changed in a sample. Procedures for deriving these formulas are discussed and some of the derivations are included in this report.

Most people know these formulas exist, and they are not, for the most part, difficult to derive. However, they are often useful and it is usually difficult to find them in print. To illustrate this point, a total of eighty-six statistics, design of experiments, probability, sampling, and quality control texts were reviewed and of that number, only two* included a discussion

*(1) Sampling Inspection by Variables, Bowker and Goode, pp. 62, 63, and 92.
(2) Techniques of Statistical Analysis, Eisenhart, Hastay, and Wallis, pp. 42-43.

on how to determine the total variance from subsample statistics. From this, it appears that while the formulas and procedures which are presented here may be well known, few authors seem to have bothered to put them in print. Forthermore, it has been observed that many people have needed certain of these formulas and not being able to locate them in print have found it necessary either to spend considerable time deriving them or simply to do without.

One obvious method for obtaining the mean and variance for a total sample is to gather the raw data from all the subsamples and compute these statistics by conventional procedures. It is equally clear that use of raw data will be unsatisfactory if the subsamples are quite large bacause of the amount of work involved; and the raw data certainly cannot be used in those frequent cases in which it is no longer available.

II. APPLICATIONS AND PRECAUTIONS. The following are uses of the procedures and formulas of this section:

A. After estimating the mean and variance for a number of different populations, a research worker may want to know the mean and variance for a population composed of a combination of these populations. This would be accomplished by combining the samples from the sub-populations to obtain a total sample.

(1) An example of this would be the case of production lots. The mean and variance will be known for a sample from each lot, but an estimate of the mean and variance for the entire production may be desired. To obtain thisit would be necessary to combine the lot samples to obtain a total sample.

(2) A second example: After conducting an analysis of variance to determine the effect of certain treatments, the research worker may want to estimate the mean and variance for a population composed of several subpopulations, each identified by a certain treatment level. For this, subsamples could be combined to form a total population.

B. Sample data may come from many sources, for example, from several parts of the country, from several agencies, or from several periods of time, and it may frequently be desirable to combine the data to form one total sample. Obviously, it may be that only the mean, variance and sample size for each subsample are available or can easily be transmitted rather than the complete raw data.

C. Frequently, sample data has been completely analyzed when it becomes evident that a few observations must be added, certain observations should be deleted, or a few should be corrected. The procedures of this report may be very useful in changing or correcting the original estimates of the mean and variance as a result of changing or correcting the basic data.

In this connection, these formulas may be useful in computing statistics associated with moving averages.

D. As a final application, those who teach statistics at the Sophomore or Junior level might find the derivation and application of some of these formulas an interesting assignment.

The main precaution to observe when using these formulas is that the total sample may represent a population with such strange or unknown characteristics that an estimate of the variance would be useless when obtained. For example, a total population composed of k normal sub-populations, each with a different mean and variance, is not likely to be normal or even close to normal.

On the other hand, it is quite possible that the characteristics of the total population will be known and the estimates of its parameters useable. For example, the sub-populations may not be normal, but it may be possible to combine them to form a normal total population. Similarly, the variance for the total population may be needed to describe the distribution of sample means, and this distribution should approach normality regardless of the distribution of the total population.

Another precaution is that one should observe whether the ratio of each subsample size to the total sample size is about the same as the ratio of the corresponding sub-population. If this is not the case, weighting factors should be introduced to obtain the correct ratios.

As a final precaution, before combining subsamples to form a total sample, one should always observe whether it is inherently reasonable to combine such data. That is to say, the subsamples may contain such different types of observations that combining them would be nonsense.

The actual differences between a total estimate and a pooled estimate of the variance should be discussed at this point.

A total variance is the variance of one complete sample, which has been broken down into two or more subsamples. No assumptions are made concerning the populations corresponding to each subsample. More specifically, no assumption is made concerning the variances of these populations. However, it is assumed that when the total variance has been obtained, its corresponding total sample corresponds to a population with known characteristics. If this were not so, there would be little purpose in a total variance.

The pooled estimate of the variance can be obtained from subsample statistics, just as a total variance. It differs, however, in that it is in no way related to a total sample or a total population. Therefore, no assumptions need be made concerning a total population. The assumption is made, however, that all subsamples come from the same population, or at least from populations which have equal variances. The pooled estimate is then an improvement over each of the estimates obtained from any single subsample.

III. DEFINITIONS.

A. k = Number of subsamples.

B. n_i = Size of the i^{th} subsample ($i = 1, 2, \ldots k$).
(If all n_i are equal, use n)

C. N = Size of the complete sample.

(1) $N = \Sigma n_i$ ($i = 1, 2, \ldots k$).

(2) $N = kn$ if all n_i are equal.

D. $\bar{x}_i$ = Mean for the i^{th} subsample.

E. $\bar{\bar{x}}$ = Mean for the total sample.

F. s^2 = Variance for the complete sample.
s = Standard deviation for the total sample.

G. s_i^2 = Variance for the i^{th} subsample.

H. s_p^2 = Pooled estimate of the variance.

IV. PROCEDURES.

A. The overall mean $\bar{\bar{x}}$:

(1) If the n_i are unequal:

$$\bar{\bar{x}} = \frac{\Sigma n_i \cdot \bar{x}_i}{N} . \qquad \text{Formula (I)}$$

(2) If the n_i are equal:

$$\bar{\bar{x}} = \frac{n\Sigma \bar{x}_i}{N} = \frac{\Sigma \bar{x}_i}{k} . \qquad \text{(II)}$$

B. Pooled Estimate of the Variance s_p^2.

The pooled estimate of the variance is actually an average of the subsample variances, and should be computed only if there is reasona e assurance that all subsamples were selected from populations with equal variances.

(1) If the n_i are unequal:

$$s_p^2 = \Sigma (n_i - 1) \cdot \frac{s_i^2}{N-k} \qquad \text{(III)}$$

(2) If the n_i are all equal:

$$s_p^2 = \frac{(n-1)\Sigma s_i^2}{N-k} = \frac{\Sigma s_i^2}{k} . \qquad \text{(IIIa)}$$

C. Determining the Variance from an Analysis of Variance Table.

One method for computing both the total variance and the pooled estimate of variance is by preparing a single variable analysis of variance table. In addition to determining the variances, it will also be possible to test the null hypothesis of the equality of subsample means. This is described by Table 1.

TABLE 1 - The Analysis of Variance Method

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	k-1	TR	tr	F
Error	N-k	E	s_p^2	
Total	N-1	T	s^2	

Table 1 is completed as follows:

(1) Complete all entries under degrees of freedom.

(2) Compute and enter:

$E = \Sigma(n_i-1)s_i^2$, or if all n_i are equal $E = (n-1)\Sigma s_i^2$.

(3) $TR = \Sigma n_i \cdot \bar{x}_i^2 - N\bar{\bar{x}}^2$, or if all n_i are equal $TR = n\Sigma\bar{x}_i^2 - N\bar{\bar{x}}^2$

$= n(\Sigma\bar{x}_i^2 - k\bar{\bar{x}}^2)$.

(4) $T = E + TR$.

(5) $s^2 = \frac{T}{N-1}$ <u>This is the desired solution.</u>

(6) The pooled estimate $s_p^2 = \frac{E}{N-k}$.

(7) If it is desired to test the null hypothesis for equality of means:

$tr = \frac{TR}{k-1}$, and

$F = \frac{tr}{s_p^2}$ with (k-1) and (N-k) degrees of freedom.

D. Formulas for the Variance of a Total Sample.

It is simple to obtain the desired formulas for the variance (and standard deviation) for the total sample by following the procedures of the analysis of variance given in the previous section. These formulas are given below:

(1) The general formula:

$$s^2 = \frac{\Sigma(n_i-1)s_i^{\,2} + \Sigma n_i \cdot \bar{x}_i^{\,2} - N\bar{\bar{x}}^2}{N-1}. \qquad \text{(IV)}$$

(2) If all n_i are equal:

$$s^2 = \frac{(n-1)\Sigma s_i^{\,2} + n(\Sigma \bar{x}_i^{\,2} - k\bar{\bar{x}}^2)}{N-1}. \qquad \text{(V)}$$

(3) If a pooled estimate of the variance is available and the n_i are unequal, formula IV may be written thus:

$$s^2 = \frac{(N-k)s_p^{\,2} + \Sigma n_i \cdot \bar{x}_i^{\,2} - N\bar{\bar{x}}^2}{N-1}. \qquad \text{(VI)}$$

(4) If a pooled estimate of the variance is available and the n_i are all equal, formula V may be written thus:

$$s^2 = \frac{(N-k)s_p^{\,2} + n(\Sigma \bar{x}_i^{\,2} - k\bar{\bar{x}}^2)}{N-1}. \qquad \text{(VII)}$$

(5) If $k = 2$ and $n_1 = n_2$, formula V may be further simplified:

$$s^2 = \frac{N-2}{2(N-1)} \cdot (s_1^{\,2} + s_2^{\,2}) + \frac{N}{4(N-1)} \cdot (\bar{x}_1 - \bar{x}_2)^2. \qquad \text{(VIII)}$$

(6) If all n_i are equal and n is large, the following approximation may be used for formula V:

$$s^2 \cong s_A^2 = \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 = s_p^2 + \frac{\Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2. \qquad \text{(IX)}$$

In Appendix II, it is shown that the error in formula IX is as follows:

$$\text{Error} = s_A^2 - s^2 = \frac{1}{N} \cdot (\Sigma s_i^2 - s^2) . \qquad \text{(X)}$$

The error described in formula X is always positive.

(7) Formula XI is offered as a substitute for formula IV and formula XII as a substitute for formula V. Actually, formulas XI and XII may require more labor than the original formulas, but they will usually involve smaller numbers and may frequently result in greater accuracy.

$$s^2 = \frac{\Sigma (n_i - 1) s_i^2 + \Sigma n_i (\bar{x}_i - \bar{\bar{x}})^2}{N-1} \qquad \text{(XI)}$$

$$s^2 = \frac{(n-1) \Sigma s_i^2 + n \Sigma (\bar{x}_i - \bar{\bar{x}})^2}{N-1} . \qquad \text{(XII)}$$

E. Formulas Associated with Changes in Data.

(1) Frequently, after computing the desired statistics for a sample of size n_1, the worker is faced with the necessity of adding an extra group of n_2 observations to the sample. If n_1 is large and n_2 relatively small, it would appear to be desirable to compute the mean and variance for the n_2 additional observations and determine the statistics for the entire sample from formulas I and IV. This technique is illustrated in Appendix I, Section D.

(2) In the event only one new observation (y) has been added to the sample, formulas XIII and XIV offer a simple procedure for obtaining the desired mean and variance. Similarly, formulas XV and XVI may be used if two observations, (y) and (w) are to be added.

$$\bar{\bar{x}} = \frac{n_1 \cdot \bar{x}_1 + y}{n_1 + 1} \qquad \text{(XIII)}$$

$$s^2 = \frac{n_1 - 1}{n_1} \cdot s_1^2 + \frac{(\bar{x}_1 - y)^2}{n_1 + 1} \qquad \text{(XIV)}$$

$$\bar{\bar{x}} = \frac{n_1 \cdot \bar{x}_1 + y + w}{n_1 + 2} \qquad \text{(XV)}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + y^2 + w^2 + n_1 \cdot \bar{x}_1^2 - (n_1 + 2)\bar{\bar{x}}^2}{n_1 + 1} \qquad \text{(XVI)}$$

(3) Similarly, after computing the mean and variance for a sample of size n_1, it may be necessary to discard n_2 observations. If the mean and variance are computed for the n_2 observations which have been discarded, formulas XVII and XVIII may be used to obtain the mean and variance for the remaining observations.

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 - n_2 \bar{x}_2}{n_1 - n_2} \qquad \text{(XVII)}$$

$$s^2 = \frac{(n_1 - 1)s_1^2 - (n_2 - 1)s_2^2 - (n_1 - n_2)\bar{\bar{x}}^2 - n_2 \bar{x}_2 + n_1 \bar{x}_1^2}{(n_1 - n_2 - 1)} \qquad \text{(XVIII)}$$

This is illustrated in section E of Appendix I.

(4) Discarding One Term

If there is only one term (y) to be discarded, formulas XIX and XX may be used.

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y}{n_1 - 1} \qquad \text{(XIX)}$$

$$s^2 = \frac{n_1 - 1}{n_1 - 2} \cdot s_1^{\,2} - \frac{n_1 \cdot (\bar{x}_1 - y)^2}{(n_1 - 1)(n_1 - 2)} \ . \qquad \text{(XX)}$$

(5) Replacing Observations

If a group of n_2 observations in a sample of size n_1 should by changed, one may follow the steps discussed in sections (1) and (3). If it is only one observation, formulas XXI, XXII and XXIII may be used. Assume y is the value to be removed and replaced by w.

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y + w}{n_1} \qquad \text{(XXI)}$$

$$s^2 = s_1^{\,2} + \frac{w^2 - y^2 - n_1(\bar{\bar{x}}^2 - \bar{x}_1^{\,2})}{n_1 - 1} \qquad \text{(XXII)}$$

or:

$$s^2 = s_1^{\,2} + \frac{(w-y)\cdot[(n_1 - 1)w + (n_1 + 1)y - 2n_1 \cdot \bar{x}_1]}{(n_1)(n_1 - 1)} \ . \qquad \text{(XXIII)}$$

F. Variance and Mean for a Total Population Composed of k Normal Populations

It appears appropriate to conclude with a brief discussion of population parameters. Assume a total population is composed of k normal sub populations, with mean μ_i and variance σ_i; and each contributing to the total population in the proportion f_i, with $\Sigma f_i = 1$. Formula XXIV gives the mean (μ) for the total population and formula XXV for the variance (σ^2) of the total population.

$$\mu = f_1\mu_1 + f_2\mu_2 + \ldots + f_k\mu_k \qquad \text{(XXIV)}$$

$$\sigma^2 = \sum_1^k t_i \sigma_i^2 + \sum_1^k t_i \cdot (\mu_i - \mu)^2, \qquad \text{(XXV)}$$

The derivation of these formulas is given in Appendix III. The chief reason for including this section is to point out the similarity between formulas XXIV and I and between XXV and XI, which is just as would be expected.

APPENDIX I - Examples

A. Example One - (All n_i equal)

(1) Consider the example given by Table 2 in which there are four equal subsamples, each of size ten.

TABLE 2

	SS(1)	SS(2)	SS(3)	SS(4)
	350	300	300	300
	340	295	310	275
	335	310	340	280
	345	315	330	310
	300	305	290	305
	325	325	285	290
	330	285	295	260
	335	310	300	325
	325	325	305	290
	335	330	290	280
n	10	10	10	10
$\bar{x}_i$	334.00	310.00	340.50	291.50
s_i^2	243.33	205.56	319.17	361.39
				N = 40
				$\bar{\bar{X}}$ = 310

(2) Using the raw data in this example, the value $s^2 = 503.85$ may be easily computed. However, it is the purpose of this example to demonstrate techniques for obtaining s^2 if the raw data is unavailable or if N is so large that it would not be feasible to use the raw data. The first step will be to use formula II to obtain $\bar{\bar{x}}$.

$$\bar{\bar{x}} = \frac{\Sigma \bar{x}_i}{k} = \frac{1240}{4} = 310.$$

(3) Table 3 demonstrates the application of the analysis of variance, as described by Table 1, to obtain s^2.

TABLE 3

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	$(k-1) = 3$	$TR = 9,485$	$tr = 3161.67$	11.20
Error	$(N-k) = 36$	$E = 10,165$	$s_p^2 = 282.36$	
TOTAL	$(N-1) = 39$	$T = 19,650$	$s^2 = 503.85$	

Where:

$$E = (n-1)\ (\Sigma s_i^2) = 10,165$$

$$TR = (n)\ (\Sigma \bar{x}_i^2) - N\bar{\bar{x}}^2 = 9,485$$

$$T = E + TR = 19,650$$

$$s^2 = \frac{T}{(N-1)} = 503.85$$

$$s = \sqrt{503.85} = 22.45.$$

If a pooled estimate of variance is desired:

$$s_p^2 = \frac{E}{(N-k)} = 282.36.$$

If it is desired to test for the equality of means:

$$tr = \frac{TR}{(k-1)} = 3{,}161.67$$

$$F = \frac{tr}{s_p^2} = 11.20$$ with 3 and 36 degrees of freedom. The value of F indicates that the difference in means is highly significant.

(4) Applying the formulas from Section III-D, one obtains:

(a) $$s^2 = \frac{(n-1)\Sigma s_i^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} = \qquad \text{(V)}$$

$$\frac{(9)(1129.45) + (10)(385{,}348.50 - 4 \times 96{,}100)}{39} = 503.85$$

(b) If a pooled estimate of the variance is available, one may use formula VII.

$$s^2 = \frac{(N-k)s_p^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} = \qquad \text{(VII)}$$

$$\frac{(36)(282.36) + (10)(385{,}348.50 - 4 \times 96{,}100)}{39} = 503.85.$$

(c) If it is desired that the numbers be kept smaller, formula XII may be used.

$$s^2 = \frac{(n-1)\Sigma s_i^2 + n\Sigma(\bar{x}_i - \bar{\bar{x}})^2}{N-1} = \qquad \text{(XII)}$$

$$\frac{(9)(1129.45) + (10)(948.50)}{39} = 503.85.$$

(d) If an approximation is desired, one may use formula IX.

$$s_A^2 = \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 =$$

$$\frac{1129.45 + 385,348.50}{4} - 96,100 = 519.49 \;;$$

giving a positive error of 15.64, exactly what formula X would indicate the error to be.

B. Example Two - (n_i unequal)

(1) Consider the following example in which there are four subsamples and a total sample size of 32; Table 4.

TABLE 4

	SS(1)	SS(2)	SS(3)	SS(4)
	350	300	300	300
	340	295	310	275
	335	310	340	280
	345	315	330	310
	355	305	290	305
	300	325	285	290
	325	285	300	
	330	310		
	325	325		
		330		
n_i	9	10	7	6
$\bar{x}_i$	333.89	310.00	307.86	293.33
s_i^2	273.61	205.56	415.48	196.67
				N = 32
				$\bar{\bar{x}}$ = 313.125

(d) If an approximation is desired, one may use formula IX.

$$s_A^2 = \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 =$$

$$\frac{1129.45 + 385,348.50}{4} - 96,100 = 519.49 \quad ;$$

giving a positive error of 15.64, exactly what formula X would indicate the error to be.

B. Example Two - (n_i unequal)

(1) Consider the following example in which there are four subsamples and a total sample size of 32; Table 4.

TABLE 4

	SS(1)	SS(2)	SS(3)	SS(4)
	350	300	300	300
	340	295	310	275
	335	310	340	280
	345	315	330	310
	355	305	290	305
	300	325	285	290
	325	285	300	
	330	310		
	325	325		
		330		
n_i	9	10	7	6
$\bar{x}_i$	333.89	310.00	307.86	293.33
s_i^2	273.61	205.56	415.48	196.67
				N = 32
				$\bar{\bar{x}}$ = 313.125

(2) From the sample of 32, the value $s^2 = 452.82$ can easily be computed.

(3) Table 5 gives the analysis of variance.

TABLE 5

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	$(k-1) = 3$	TR= 6529.68	tr=2176.56	8.11
Error	$(N-k) = 28$	E= 7515.15	$s_p^2 = 268.40$	
TOTAL	$(N-1) = 31$	T=14,044.83	$s^2 = 453.06$	

Where:

$$E = \Sigma(n_i-1)\, s_i^2 = 7515.15$$

$$TR = \Sigma n_i \bar{x}_i^2 - N\bar{\bar{x}}^2 = 3{,}144{,}042.18 - 3{,}137{,}512.50 = 6529.68$$

$$T = E + TR = 14{,}044.83$$

$$s^2 = \frac{T}{(N-1)} = 453.06.$$

(Note that there is a slight difference between this estimate and the one obtained from the basic data, due to rounding errors.)

$$s = \sqrt{453.06} = 21.29.$$

If a pooled estimate of variance is desired:

$$s_p^2 = \frac{E}{(N-k)} = 268.40.$$

If it is desired to test the equality of the means for the four subsamples:

$$tr = \frac{TR}{(k-1)} = 2176.56$$

$$F = \frac{tr}{s_p^2} = 8.11$$ with 3 and 28 degrees of freedom. This indicates that the difference in means is highly significant.

(4) Applying the formulas from Section III-D, one obtains:

(a) $$s^2 = \frac{\Sigma(n_i-1)s_i^2 + \Sigma n_i\bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} = \qquad \text{(IV)}$$

$$\frac{7515.15 + 3,144,042.18 - 3,137,512.50}{31} = 453.06,$$

(b) If a pooled estimate of the variance is available, formula VI may be used.

$$s^2 = \frac{(N-k)s_p^2 + \Sigma n_i\bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} = \qquad \text{(VI)}$$

$$\frac{(28)(268.40) + 3,144,042.18 - 3,137,512.50}{31} = 453.11.$$

(c) If it is desired that the numbers be kept small, formula XI may be used.

$$s^2 = \frac{\Sigma(n_i-1)s_i^2 + \Sigma n_i(\bar{x}_i - \bar{\bar{x}})^2}{N-1} = \qquad \text{(XI)}$$

$$\frac{7515.15 + 6523.42}{31} = 452.86.$$

(Note that this is much closer to the true value than those listed under a or b).

C. Example Three - ($k = 2, n_1 = n_2$)

(1) To illustrate formula VIII, the first two columns from Table 2 will be used. From this:

$n = 10, \; N = 20$

$\bar{x}_1 = 334, \; \bar{x}_2 = 310$

$s_1^2 = 243.33, \; s_2^2 = 205.56$

$\bar{\bar{x}} = 322, \; s_p^2 = 224.44, \; s^2 = 364.21.$

(2) Applying formula VIII:

$$s^2 = \frac{(N-2)}{2(N-1)} \cdot (s_1^2 + s_2^2) + \frac{N}{4(N-1)} \cdot (\bar{x}_1 - \bar{x}_2)^2 =$$

$$\frac{9}{19}(448.89) + \frac{5}{19}(576) = 364.24.$$

D. Example Four: (Add n_2 observations to a sample of size n_1)

(1) Consider the sample of 40, given by Table 2. It may be observed that:

$\bar{x}_1 = 310, \; s_1^2 = 503.85.$

(2) Suppose it is necessary to add the five additional items:

310, 293, 314, 280, and 300

$n_2 = 5, \; \bar{x}_2 = 305.40, \; s_2^2 = 374.80.$

(3) One may proceed by using formula I and then either IV or XI. For this formula, XI was used.

$N = 40 + 5 = 45$

$$\bar{\bar{x}} = \frac{(40)(310.00) + (5)(305.40)}{45} = 309.49$$

by formula I.

$$s^2 = \frac{(39)(503.85) + (4)(374.80) + (40)(.51)^2 + (4)(4.09)^2}{44}$$

$$= 482.42$$

by formula XI.

(4) Actually, the value for s^2 using raw data is 482.80.

E. Example Five: (Remove n_2 observations from a sample of size n_1.)

(1) The data of example four will be used for this.

$n_1 = 45$

$\bar{x}_1 = 309.49$

$s_1^2 = 482.80.$

(2) Remove the 5 observations which were added in example four.

$n_2 = 5$

$\bar{x}_2 = 305.40$

$s_2^2 = 374.80.$

(3) Use formulas XVII and XVIII, giving:

$$\bar{\bar{x}} = \frac{(45)(309.49) - (5)(305.40)}{40} = 310.00$$

$$s^2 = \frac{(44)(482.80) - (4)(374.80) - (40)(310)^2 - (5)(305.40)^2 + (45)(309.49)^2}{39}$$

$$= 504.64.$$

APPENDIX II - DETERMINATION OF THE ERROR IN FORMULA IX

Using formulas V and IX, the following error is observed:

$$E = s_A^2 - s^2 = \left\{ \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 \right\} - \frac{(n-1)\Sigma s_i^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1}$$

$$= \frac{(N-n)\Sigma s_i^2}{N(N-1)} - \frac{n\Sigma \bar{x}_i^2}{N(N-1)} + \frac{\bar{\bar{x}}^2}{N-1}$$

$$= \frac{\Sigma s_i^2}{N} - \frac{(n-1)\Sigma s_i^2}{N(N-1)} - \frac{n\Sigma \bar{x}_i^2}{N(N-1)} + \frac{kn\bar{\bar{x}}^2}{N(N-1)}$$

$$\text{Error} = s_A^2 - s^2 = \frac{1}{N} \cdot (\Sigma s_i^2 - s^2).$$

Inasmuch as Σs_i^2 is larger than s^2, the error will always be on the positive side.

APPENDIX III - DERIVATION OF THE MEAN AND VARIANCE FOR A POPULATION COMPOSED OF k NORMAL POPULATIONS

A. Assume each of the k normal populations have a mean μ_i, variance σ_i^2, and contributes to the total population in the proportion f_i, with $f_1 + f_2 + \ldots + f_k = 1$.

B. $$y = \frac{f_1}{\sqrt{2\pi}\sigma_1} \cdot e^{\left(-\frac{1}{2\sigma_1^2}(x-\mu_1)^2\right)} + \ldots + \frac{f_k}{\sqrt{2\pi}\sigma_k} \cdot e^{\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)}.$$

C. $$m(\theta) = f_1 e^{\left(\frac{1}{2} \cdot \theta^2\sigma_1^2 + \theta\mu_1\right)} + \ldots + f_k e^{\left(\frac{1}{2} \cdot \theta^2\sigma_k^2 + \theta\mu_k\right)}.$$

D. $$\frac{\partial m(\theta)}{\partial\theta} = f_1(\theta\sigma_1^{\ 2} + \mu_1)\cdot e^{\left(\frac{1}{2}\cdot\theta^2\sigma_1^{\ 2} + \theta\mu_1\right)} + \dots + f_k(\theta\sigma_k^{\ 2} + \mu_k)\cdot e^{\left(\frac{1}{2}\cdot\theta^2\sigma_k^{\ 2} + \theta\mu_k\right)},$$

E. The mean of the total population:

$$\mu = f_1\mu_1 + f_2\mu_2 + \dots + f_k\mu_k.$$

F. $$\frac{\partial m_{(y-\mu)}\cdot\theta}{\partial\theta} = f_1\left[\theta\sigma_1^{\ 2} + (\mu_1-\mu)\right]\cdot e^{\left(\frac{1}{2}\theta^2\sigma_1^{\ 2} + \theta(\mu_1-\mu)\right)} + \dots,$$

G. $$\frac{\partial^2 m_{(y-\mu)}\cdot\theta}{\partial^2\theta} = f_1\left[\theta\sigma_1^{\ 2} + (\mu_1-\mu)\right]^2\cdot e^{\left(\frac{1}{2}\theta^2\sigma_1^{\ 2} + \theta(\mu_1-\mu)\right)} + f_1\sigma_1^{\ 2}\cdot e^{\left(\frac{1}{2}\theta^2\sigma_1^{\ 2} + \theta(\mu_1-\mu)\right)} + \dots,$$

H. The variance for the total population:

$$\sigma_y^{\ 2} = f_1(\mu_1-\mu)^2 + f_1\sigma_1^{\ 2} + \dots + f_k(\mu_k-\mu)^2 + f_k\sigma_k^{\ 2} = \Sigma f_i\sigma_i^{\ 2} + \Sigma f_i(\mu_i-\mu)^2.$$

APPENDIX IV - SUMMARY OF FORMULAS

A. $n_1 = n_2 = \dots n_k$

(1) The Mean for the total sample

$$\bar{\bar{x}} = \frac{\Sigma\bar{x}_i}{k}. \qquad \text{Formula II}$$

(2) Pooled Estimate of the Variance

$$s_p^2 = \frac{\Sigma s_i^2}{k}. \qquad \text{Formula III (a)}$$

(3) Variance for the total sample

$$s^2 = \frac{(n-1)\Sigma s_i^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} \qquad \text{Formula V}$$

$$s^2 = \frac{(N-k)s_p^2 + n(\Sigma \bar{x}_i^2 - k\bar{\bar{x}}^2)}{N-1} \qquad \text{Formula VII}$$

$$s^2 = \frac{(n-1)\ \Sigma s_i^2 \ + \ n\Sigma(\bar{x}_i - \bar{\bar{x}})^2}{N-1}, \qquad \text{Formula XII}$$

(4) Approximation (Use only if n is large.)

$$s_A^2 = \frac{\Sigma s_i^2 + \Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2 = s_p^2 + \frac{\Sigma \bar{x}_i^2}{k} - \bar{\bar{x}}^2. \qquad \text{Formula IX}$$

The error in Formula IX

$$\text{Error} = s_A^2 - s^2 = \frac{1}{N} \cdot (\Sigma s_i^2 - s^2) \geq 0. \qquad \text{Formula X}$$

(5) $k = 2,\ n_1 = n_2$

$$s^2 = \frac{N-2}{2(n-1)} \cdot (s_1^2 + s_2^2) + \frac{N}{4(N-1)} \cdot (\bar{x}_1 - \bar{x}_2)^2. \qquad \text{Formula VIII}$$

B. The n_i are unequal

(1) The mean for the total sample

$$\bar{\bar{x}} = \frac{\Sigma n_i \bar{x}_i}{N}. \qquad \text{Formula I}$$

(2) Pooled estimate of the variance

$$s_p^2 = \frac{\Sigma (n_i - 1) s_i^2}{N-k} \qquad \text{Formula III}$$

(3) Variance for the total sample

$$s^2 = \frac{\Sigma (n_i - 1) s_i^2 + \Sigma n_i \bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} \qquad \text{Formula IV}$$

$$s^2 = \frac{(N-k) s_p^2 + \Sigma n_i \bar{x}_i^2 - N\bar{\bar{x}}^2}{N-1} \qquad \text{Formula VI}$$

$$s^2 = \frac{\Sigma (n_i - 1) s_i^2 + \Sigma n_i (\bar{x}_i - \bar{\bar{x}})^2}{N-1} \qquad \text{Formula XI}$$

C. Formulas associated with changes in data

(1) Add an observation y to a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 + y}{n_1 + 1} \qquad \text{Formula XIII}$$

$$s^2 = \frac{n_1 - 1}{n_1} \cdot s_1^2 + \frac{(\bar{x}_1 - y)^2}{n_1 + 1} \qquad \text{Formula XIV}$$

(2) Add observations y and w to a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1 \bar{x}_1 + y + w}{n_1 + 2} \qquad \text{Formula XV}$$

$$s^2 = \frac{(n_1 - 1) s_1^2 + y^2 + w^2 + n_1 \bar{x}_1^2 - (n_1 + 2)\bar{\bar{x}}^2}{n_1 + 1} \qquad \text{Formula XVI}$$

(3) Discard n_2 observations from a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - n_2\bar{x}_2}{n_1 - n_2} \qquad \text{Formula XVII}$$

$$s^2 = \frac{(n_1-1)s_1^{\;2} - (n_2-1)s_2^{\;2} + n_1(\bar{x}_1^{\;2}-\bar{\bar{x}}^2) - n_2(\bar{x}_2^{\;2}-\bar{\bar{x}}^2)}{(n_1 - n_2 - 1)}, \qquad \text{Formula XVIII}$$

(4) Discard the observation (y) from a sample of size n_1

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y}{n_1 - 1} \qquad \text{Formula XIX}$$

$$s^2 = \frac{n_1-1}{n_1-2} \cdot s_1^{\;2} - \frac{n_1(\bar{x}_1-y)^2}{(n_1-1)(n_1-2)} . \qquad \text{Formula XX}$$

(5) Replace the observation y by w

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 - y + w}{n_1} \qquad \text{Formula XXI}$$

$$s^2 = s_1^{\;2} + \frac{w^2-y^2-n_1(\bar{\bar{x}}^2-\bar{x}_1^{\;2})}{n_1 - 1} \qquad \text{Formula XXII}$$

or

$$s^2 = s_1^{\;2} + \frac{(w-y)\left[(n_1-1)w + (n_1+1)y - 2n_1\bar{x}_1\right]}{(n_1)\,(n_1-1)} . \qquad \text{Formula XXIII}$$

D. Formulas associated with a total population composed of k normal populations

$$\text{Let } y = \frac{f_1}{\sqrt{2\pi}\sigma_1} \cdot e^{\left(-\frac{1}{2\sigma_1^{\;2}} \cdot (x-\mu_1)^2\right)} + \ldots + \frac{f_k}{\sqrt{2\pi}\sigma_k} \cdot e^{\left(-\frac{1}{2\sigma_k^{\;2}} \cdot (x-\mu_k)^2\right)}$$

Where $f_1 + f_2 + \ldots + f_k = 1$, then:

$$\mu = f_1\mu_1 + f_2\mu_2 + \ldots + f_k\mu_k$$ Formula XXIV

$$\sigma^2 = \Sigma f_i \sigma_i^2 + \Sigma f_i(\mu_i - \mu)^2.$$ Formula XXV

SYSTEM CONFIGURATION PROBLEMS AND ERROR SEPARATION PROBLEMS*

Fred S. Hanson
Plans and Operations Directorate
White Sands Missile Range, New Mexico

ABSTRACT. Practical geometric criteria and optimization methods are needed for laying out, or selecting, multi-instrument configurations for flight measurement. The problem is to discover - and demonstrate - some principles that are at least in the right direction. A general solution should be possible for the variation of uncertainty of intersection location as a function of angles-of-intersection of lines-of-sight. It might also be possible to calculate the optimum ground-pattern for a given station density and missile trajectory. The second problem is to develop - in detail - analytical tools for separating position-measurement error, time-measurement error, and lack-of-fit of a given polynomial -- as these errors exist in undesigned, but redundant, data. Questions concern: the validity of linearization of data for this purpose; procedures for calculating lack-of-fit of polynomials of degrees greater than one; limitations in conversion of regressions to analyses of variance.

INTRODUCTION. This paper is clinical -- especially in the sense that it is not completed work.

BACKGROUND. Figure 1 is a White Sands Missile Range briefing chart. It shows: the principal Range (heavy line); the part-time extension (at the top); and the White Sands Monument (small internal area). Headquarters - and the main launch areas - are at the lower end of the Range.

The distinction between optical and electronic tracking instruments has been lost in this black-and-white print. Optical instruments include: cinetheodolites, telescopes, fixed cameras, and ballistic cameras. Not every station is shown. For instance, there are several hundred prepared sites where fixed cameras can be set up. Electronic tracking instruments include: radars, dopplers, and miss-distance systems. Again, not every station is shown. (There are several hundred prepared sites where DOVAP receivers can be set up.) The gray - and part-gray - dots are telemetry receivers.

*Comments on this paper by some of the panelists can be found following the figures at the end of this article.

It may be apparent that the systems in Figure 1 were not laid out on any rigorous basis.

CONFIGURATION HYPOTHESES. More than three years ago (Ref. 1), the writer asserted two hypotheses about instrument layout, or selection -- to initiate action toward solution.

First, it was asserted - intuitively - that the most favorable elevation angle for observing a missile is 45°. Second, the writer stated an optimum ground-configuration - for each integral number of stations - with respect to a single point in space. This was done on the assumption that the best intersection of lines-of-sight from two stations is - when considered by itself - 90°. Conversely, it was assumed that the worst intersection occures when one station looks over another's shoulder, or they look down each others throats -- 0° or 180°, parallel. Referring to Figure 2, the most favorable ground-configuration for optical stations was asserted - without proof - to be: two-station - right-isosceles triangle with missile at apex; three-station - equilateral triangle with missile at center; (in all subsequent cases, missile at center) four-station - any four corners of equilateral pentagon; five-station - said pentagon; six-station - any six corners of equilateral heptagon; seven-station - that heptagon; etc. The (corresponding) intersection angles are: 90°, 120°, 72°, and 51.4°. For twelve or thirteen stations - a tridecagon - the angle would be down to 27.7°.

DEMONSTRATION OF HYPOTHESES. After proposing this paper, the writer made a crude approach to demonstrating (the validity of) these simple hypotheses.

Figure 3a shows the asserted two-station optimum. This can be any plane through both stations and the missile. The diagram represents the 90° intersection - together with some dispersion index, such as the standard deviation.

Figure 3b is an enlargement of the area of uncertainty. We are assuming the two instruments are equally precise. Let's approximate the actual error-ellipse by the almost-square in Figure 3a - and approximate that by the square in Figure 3b. The horizontal diagonal is a measure of the combined error-variance. If we increase the intersection angle, by moving the stations farther apart - or by lowering the missile - the horizontal diagonal will lengthen. Of course, the vertical diagonal will

shorten, correspondingly. In general, it's not sound practice to improve data in one coordinate by making it worse in another. (If we decrease the intersection angle - below 90° -, the horizontal diagonal gets smaller, at the expense of the vertical diagonal.) So, we may conclude 90° is the practical optimum.

Now, we have shown that 90° is the optimum intersection in any plane thru both stations and the missile. The plan for which the degradations from this optimum will be the same in its horizontal and vertical projections is the 45° plane. On the basis that there is no preferred coordinate, we have demonstrated the hypothesis regarding the optimum elevation angle.

If we choose to take our geometry in algebraic form, we can use the law of cosines to calculate the horizontal diagonal (Figure 3b):

$$a^2 = b^2 + c^2 - 2bc \cos \theta$$

where b and c are measures of the two observational variances. θ is approximately 90°. To see the effect of changing the intersection from 90°, let's replace θ by $90° \pm \alpha$:

$$a^2 = b^2 + c^2 - 2bc \cos(90° \pm \alpha)$$

In our case, b and c are equal, so:

$$a^2 = 2b^2 - 2b^2 \cos(90° \pm \alpha)$$

$$= 2b^2 \; [1 - \cos(90° \pm \alpha)]$$

Substituting, $$a^2 = 2b^2 (1 \mp \sin \alpha).$$

So, approximately, if the intersection angle is changed, the combined variance in one coordinate increases as the sine of the angular deviation from 90°.

A similar exercise can be gone thru for the 3-station equilateral triangle. In that case, the error-ellipse is approximated by an almost-equilateral hexagon.

MORE ANALYTICAL SOLUTION. W. F. Mammack (Ref. 2) has furnished the writer a solution which does not depend on approximating the almost-square -- or on testing a hypothesis.

Referring to Figure 4a - the trigonometry for the general two-station case yields:

$$x = \frac{b}{2}\,\frac{\sin(\theta_1 - \theta_2)}{\sin(\theta_1 + \theta_2)} \qquad y = b\,\frac{\sin\theta_1 \sin\theta_2}{\sin(\theta_1 + \theta_2)}$$

Applying the standard error-propagation formula:

$$\epsilon_x^2 = \left(\frac{\partial x}{\partial \theta_1}\right)\epsilon_{\theta_1}^2 + \left(\frac{\partial x}{\partial \theta_2}\right)\epsilon_{\theta_2}^2$$

(and similarly for y) yields:

$$\epsilon_x^2 = \frac{b^2}{\sin^4(\theta_1 + \theta_2)}\left(\frac{1}{4}\sin^2 2\theta_2\,\epsilon_{\theta_1}^2 + \frac{1}{4}\sin^2 2\theta_1\,\epsilon_{\theta_2}^2\right)$$

$$\epsilon_y^2 = \frac{b^2}{\sin^4(\theta_1+\theta_2)}\cdot\left(\sin^4\theta_2\,\epsilon_{\theta_1}^2 + \sin^4\theta_1\,\epsilon_{\theta_1}^2\right).$$

Simplifying to the equidistant, equal-precision case (Figure 4b):

$$\mu^2 = \epsilon_x^2 + \epsilon_y^2 = b^2\,\frac{(1 - \cos 2\theta)}{\sin^4 2\theta}.$$

If this total error is minimized with respect to θ, the minimum is found to occur at:

$$\cos 2\theta = 1/3$$

$$2\theta = 70.5^\circ.$$

So, Mimmack's optimum intersection angle is 109.5°.

In R. C. Davis' NOTS report on his cinetheodolite-reduction method (Ref. 3), he minimized the observational error-ellipse of the two-station-missile triangle, by a matrix process. With the stations fixed and the missile altitude allowed to vary, Davis found the optimum intersection to be 120°. He theorized this was the result of compromise between the most favorable intersection and the decrease in the linear error (corresponding to a given angular error) as the missile moves closer to the stations. Mimmack's solution represents this same case. So, there is an apparent discrepancy in their results.

With the missile altitude fixed and the stations free to move, Davis found the optimum intersection to be 60°. He theorized this was the result of compromise between most favorable intersection and moving the stations closer to the missile. The present writer thinks Davis' explanations are correct.

However, it appears that the optimum ground-configurations hypothesized in this paper are still optimum when the effect of slant range is included. Also, 45° planes are the only ones for which the degradations (of coordinate projections) from the optimum intersection will be the same - whatever the optimum may be. So, we have "demonstrated" a simple set of rules for laying out, or selecting, a group of stations - for any given point on a missile trajectory -, and for determining the optimum scale of their configuration. The point used could be the mid-point of a trajectory segment.

MINIMUM BIAS CONFIGURATION. The demonstration based on Figure 3 treated error as a dispersion index (or precision index). Let's consider (it as) a discrete, or net, error. Then, in Figure 3a, if we increase θ above 90°, the horizontal (error-) resultant - corresponding in size to the smaller almost-square - will lengthen if the (discrete angular) errors happen to have the same sign (Figure 5a); if the errors have opposite signs (Figure 5b), their (vertical) resultant will shorten correspondingly. (Of course - in the equal-accuracy case - there will be only a horizontal, or only a vertical, resultant.) In general, it's not sound practice to (set out to) improve data in one coordinate by taking an even chance that we will, instead, make it worse in another. (Even chance, because - to the extent that a given-type instrument consistently

has the same sign, it is more likely to be adjusted, or corrected for.) If we decrease θ (below 90°), the possible homopolar (horizontal) error-resultant gets smaller, at the expense of the possible heteropolar (vertical) error-resultant gets smaller, at the expense of the possible heteropolar (vertical) error-resultant. So, we may conclude - 90° is the practical optimum. The rest of the writer's geometric and algebraic demonstrations apply similarly. Summary: perpendicular intersection (per se), 45° elevation, the right-isosceles triangle for the two-station case, etc. are all optimum for accuracy as well as precision.

PATTERN HYPOTHESES. How does one generalize from a single group of stations to a larger area -- for (several segments of) a family of trajectories? What sort of patterns can we construct with our optimum figures? In Figure 6, what is wrong with a grid built up of optimum three-station configurations? Equilateral triangels form hexagons, which violates our odd-sided rule. Each station is in line with all the other stations. Continuing in Figure 6, pentagons seem to form a desirable pattern - leaving a few gaps of isosceles-triangle pairs. (Four stations are in line across each triangle pair.) Heptagons might do as well.

In determining the optimum layout, the decisive constraint could be the number of stations needed to meet requirements (for precision). Or, it could be budgetary (the number of stations permitted per hundred sq. mi.). Or, it could be the effective range of a station - as a configuration radius.

Perhaps someone can demonstrate that the optimum pattern is random. Or, that a random pattern is not optimum. A random pattern might have the minimum percent of stations in line with each other - but it wouldn't be the most efficient dispersion. Mimmack (Ref. 2) notes that it is desirable for a position measurement to be independent of any coordinate system; that this implies the station geometry should be free of symmetries; that the symmetry of being in the same ground-plane is largely unavoidable.

DISCUSSION OF CONFIGURATION. The optimum configuration would maximize: accuracy, precision, versatility, reliability, and economy. Flight-measuring instruments exist in three conditions: fixed, (self-contained) mobile, transportable (to prepared sites).

The writer chose to start with the precision of a single point-in-space, because this is WSMR's operating standard - and because it lends itself to an analytical approach which proceeds from the simple to the complex. The Range's instrumentation plans are prepared per segment of a trajectory. The present standard seems to be the best (single) compromise between an operating viewpoint and a missile-engineer viewpoint. Aside from having a consistent benchmark, the important question is: "What aspect of a given missile-performance variable is most significant to a particular missile project?"

This is, after all, a clinical paper. The writer's aim is not - necessarily - to solve the whole problem by an analytical approach. (It is to increase understanding of the subject.) We "demonstrated" the "90° - optimum" intersection in any plane - for observing a point-in-space. We found a (limited) approximate solution, in two dimensions, for the variation of uncertainty-of-intersection-location as a function of angle-of-intersection-of-lines-of-sight. Mimmack (Ref. 2) obtained a general solution (to this problem) for two dimensions; his method could be extended to three dimensions. It may be that an optimum ground-pattern can be constructed with pentagons.

The optimum-overall-pattern problem could be stated: "Is there a unique solution for the most efficient layout, for a given optical-station density - or for a given effective station-range - and for the Range's total trajectory-volume?" It seems clear that any thorogoing analysis of this problem must be made in three dimensions.

Reference 4, revised annually, discusses computer programs for propagating "typical" errors-of-observation thru the (trigonometric) equations relating coordinates of any given point-in-space to the (angular, etc.) "observations" of the point by stations-of-known-location. These are essentially the same programs used for trial-and-error simulation at White Sands. AMR (now ETR) calls the - a priori - error estimates so obtained "a geometric dilution of precision (GDOP)". Properly, this term should be reserved for the geometric component of position-measurement variance.

ERROR SEPARATION PROBLEM. The second problem is this: "Can we determine (by statistical methods) - qualitatively and quantitatively - how much of the error-variance in our (final) missile-position data is position-error, and how much is time-error?" For velocity and

acceleration (or smoothed position data), we would also like to know the relative magnitude of a third variance component - the lack-of-fit of the polynomial which we use to obtain (smoothed and) derivative data.

The jitter (and wander) of time-signal generators is small. Propagation- and receiver-delays are appreciable - different for each station - somewhat variable - and partly compensated for. Recording delays for: time-code marks, missile image, (angular) dial readings, etc. are appreciable, different, and somewhat variable. Overall time-measurement error includes errors in synchronizing: timing, missile position, and mount position -- physically, on the record, in conversion, in computing, and in reporting.

For a Mach 10 missile, a millisecond overall time-measurement error would be equivalent to a position error of 10 ft. A recent figure for the speed of an ICBM warhead is 26,400 ft/sec (Ref. 5); in that case a millisecond is 26.4 ft.

Actual requirements - and capabilities - for instrumentation timing-and-synchronization should be known - in specifiable terms. A complete description of position accuracy - or precision - would include a separate specification of time accuracy - or precision. If time-measurement error is ignored, it shows up as position error - but, it cannot be decreased by improving the position-measuring device (as such). If time-measurement error is appreciable, these two components of position error should be separated before calculating velocity (or acceleration) error. We don't know that time-measurement error is an appreciable part of the whole - but we can't afford not to know how much it is.

This paper presents problems -- not solutions. But - in presenting this problem - let's review the approaches the writer has already considered.

SEMI-QUANTITATIVE SEPARATION. About four years ago, the writer suggested a semi-quantitative method for "separating" time error from position error - in final data. Let's look at the three types of "regression" (correlation) of a position coordinate and time (Figure 7).

Figure 7a shows regression of x as a function of t - in which time is assumed to be exactly measured, and that curve is fitted which minimizes the (sums of the squares of) the deviations in position. This is the one WSMR uses, in its data reduction.

Figure 7b shows regression of t as a function of x - in which position is assumed to be exactly measured, and that curve is fitted which minimizes the (sums of squares of) the deviations in time. From a mathematical standpoint, this is as logical as the first.

Figure 7c shows simultaneous regression of x and t - in which they are assumed to be measured equally well, and that curve is chosen which minimizes the (ss of) the deviations. This is sometimes called the "best fit".

If measurements of x and t are about equally in error, curve c will (tend to) fall about halfway between a and b - and is the best choice, in this case.

If one variable is badly measured, the curve which minimizes the variability of the badly measured variable will (tend to) deviate the most from the other two -- but will (tend to) be closest to the (physically) true relationship. This justifies use of method a (by WSMR) - if the assumption that position is (always) much more poorly measured proves correct. The curve of "best fit" - c - best represents the data, as such, in any case.

By comparing these three types of regression - and taking into account any knowledge of the (physically) true curve from independent data, and/or physical theory -- it is possible to obtain semi-quantitative estimates of how relatively well two variables are measured. The writer knows from experience this works in applying linear regression to rather poor data. It may be an even sharper tool in applying curvilinear regression to rather good data.

QUANTITATIVE SEPARATION. On the basis of redundancy in measuring missile position, these three regressions can be converted to corresponding analyses of variance. This should permit quantitative separation of time error and position error. Procedures are available for analysis of variance of types a and b regression. Type c regression could be handled - for the linear case - by these same (single-fixed-variate) methods, by a rotation of axes. It may also be possible to discover (or devise) a bivariate analysis - at least for the linear case. If necessary curvilinear data can be transformed to linear.

Such analyses of variance include a lack-of-fit term, which is available for the linear fixed-variate case in Reference 6. It appears to be available for the curvilinear fixed-variate case from (such sources as) References 7 and 8.

The usual procedure at WSMR is to fit a second-degree polynomial. If our lack-of-fit proves to be appreciable compared to position-error, it will follow that we need to improve our data-reduction procedure.

The writer's questions with regard to the above analyses of variance are these:

1. What analysis-of-variance components can we get from linear fixed-variate regressions of types a and b if we have (apparent) redundancy in (a given) position (coordinate) at (equally-spaced) apparent times -- and (if we) convert these assumed-x redundancies to assumed-t redundancies by (means of) the reciprocal-of-the-slope of the type a regression (i.e., if we multiply by the corresponding value of $\Delta t/\Delta x$). Specifically, can we separate timing-error, position-error, and (two) lack-of-fit terms? As a working reference for this would the Panel recommend Reference 9 - or some other? Same questions for curvilinear case -- using the reciprocal of the type a slope at each point to convert - and substituting Reference 10 as a working source.

2. Suppose we apply this fixed-variate analysis to type c linear regression by a rotation of axes -- and calculate the assumed-normal redundancies by interpolating between the assumed-x and the (corresponding) assumed-t redundancies, above (in proportion to the ratio of the angle-between-the-x-axis-and-normal to 90°). Can we get anything out of this transformed type c analysis of variance?

3. Can the Panel give a reference which shows how to calculate lack-of-fit for type c linear regression?

4. Can the Panel give a reference to - or device - a bivariate analysis of variance for linear regression if we have (apparent) redundancy in (a given) position (coordinate) at (equally-spaced) apparent times? Same question for curvilinear regression.

5. Suppose we transform a variable to linearize a (curvilinear) regression -- and then perform the (linear) analysis of variance under question 1. Is it necessary to leave the result in the transformed state? Is it valid to "untransform" the variance of the transformed variable? Can the Panel give a reference on estimating the error due to "untransforming"?

6. Does Reference 7, 8, or 10 clearly give a procedure for calculating lack-of-fit for curvilinear single-fixed-variate regression? If not, can the Panel give a reference which does?

SEPARATION AT A POINT. So far we've taken a time-varying look at the flight-measurement process. White Sands is also interested in (knowing) the uncertainties associated with single values of unsmoothed data. It should be possible to make a hypothetical - if inconclusive - analysis of the errors of a single point (in space and time) by looking at the error as all (in) position, all (in) time, all tangential, or all normal. An additional approach to the "instantaneous" aspect might be to consider (two) successive data-points as observations of their mean point. Can we get any - qualitative or quantitative - separation of timing and position error out of these approaches? Can the Panel suggest any further approach to analysis of the errors of single-values-of-unsmoothed-data?

REFERENCES

1. WSMR Study Group Report, "Formulating Requirements for Range Instrumentation": DF, Supplement to Part b., 1962.

2. Letter from Wm. E. Mimmack, White Sands Missile Range - on leave to Graduate School, University of Rochester - 1965.

3. Davis, R. C. "Techniques for the Statistical Analysis of Cinetheodolite Data", NAVORD Report 1299 (NOTS 369), China Lake, Cal., 1951.

4. Mann, H. P. "The Accuracy of AMR Instrumentation", RCA Systems Analysis Technical Report 23, Patrick Air Force Base, Fla., 1962.

5. Drewry, I. O. "Hot Rod Missile", Army Information Digest, May 1965, pp. 22-56.

6. Anderson and Bancroft, Statistical Theory in Research, McGraw-Hill, 1952, pp. 156-158.

7. Schultz, H., "The Standard Error of a Forecast from a Curve", J. Amer. Stat. Assoc., 25, 139-185 (1930).

8. Ezekiel and Fox, Methods of Correlation and Regression Analysis, (3rd ed.) John Wiley & Sons, 1959.

9. Acton, Analysis of Straight-Line Data, John Wiley & Sons, 1959.

10. Brownlee, Statistical Theory and Methodology in Science and Engineering, John Wiley & Sons, 1960.

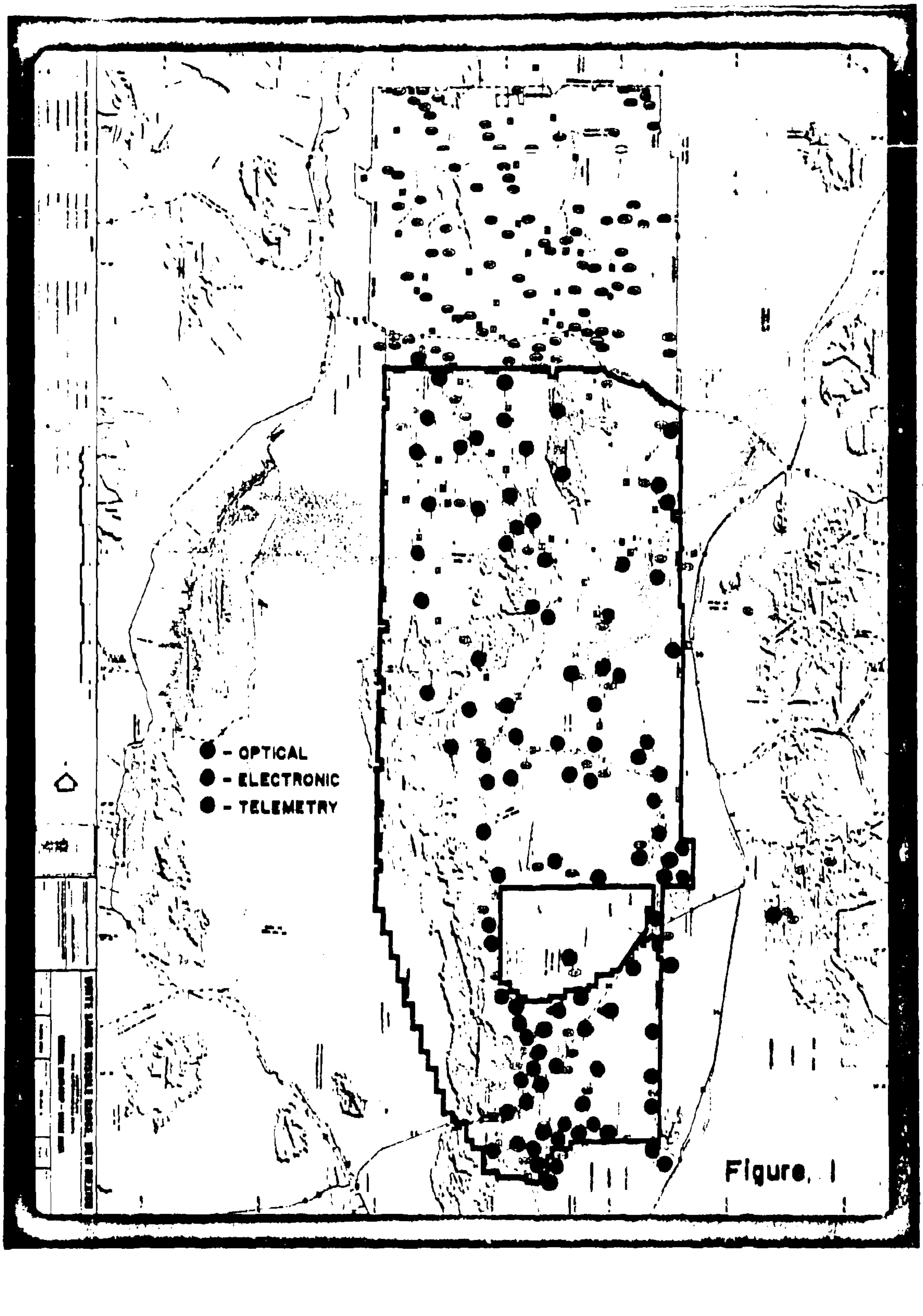


Figure. 1

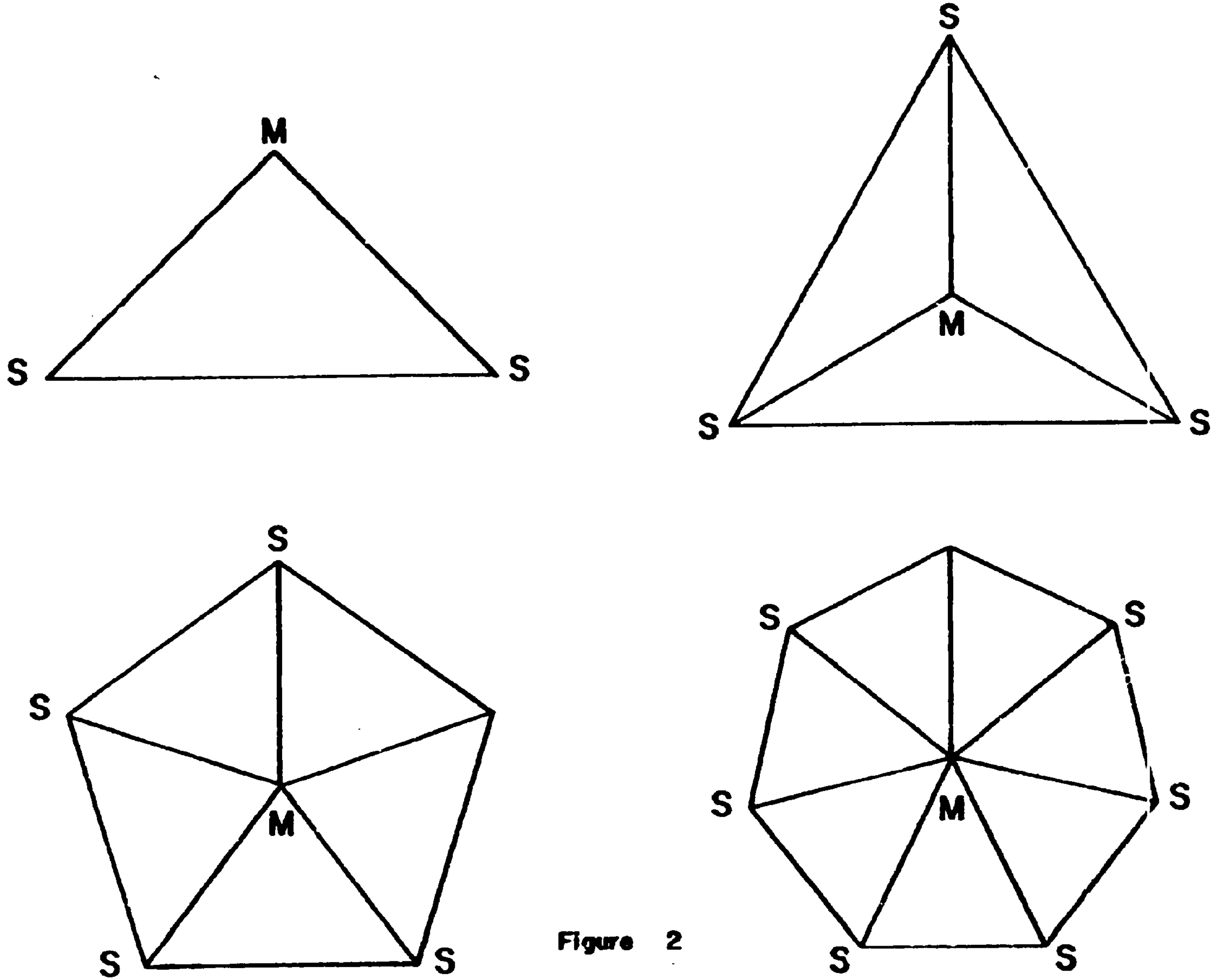


Figure 2

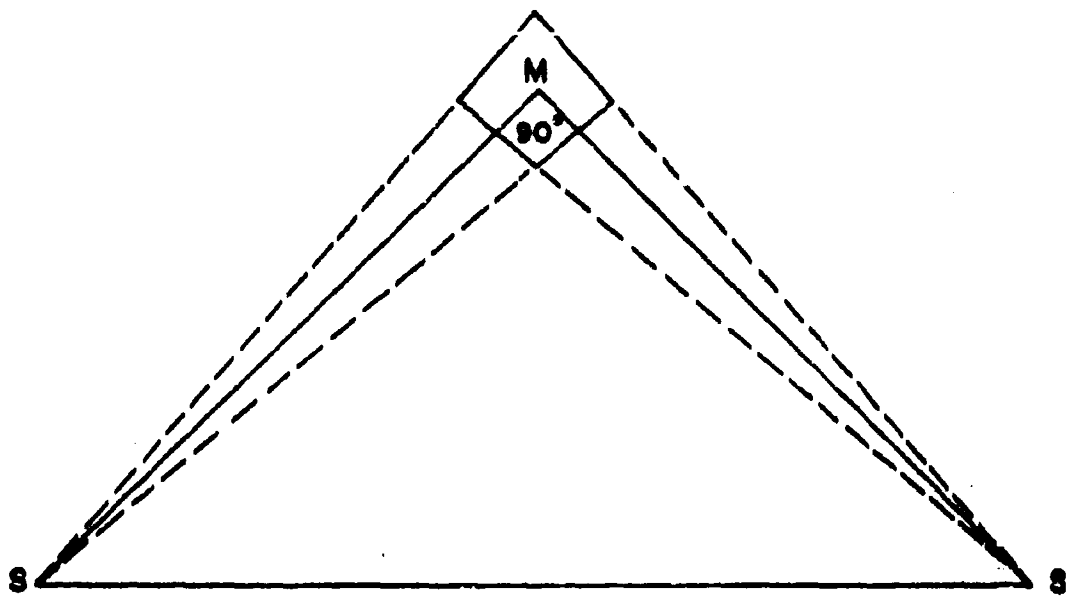


Figure 3a

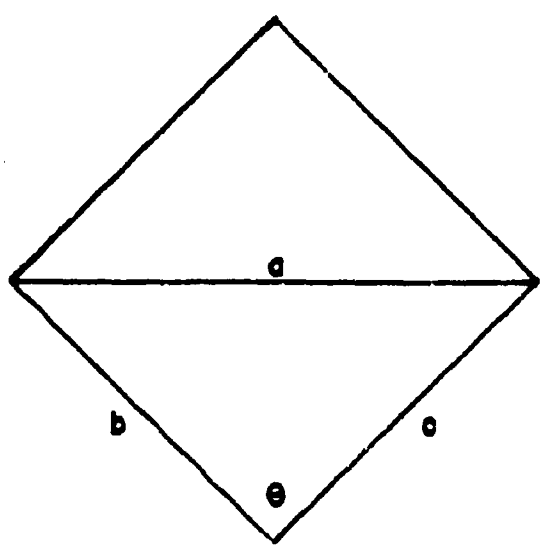


Figure 3b

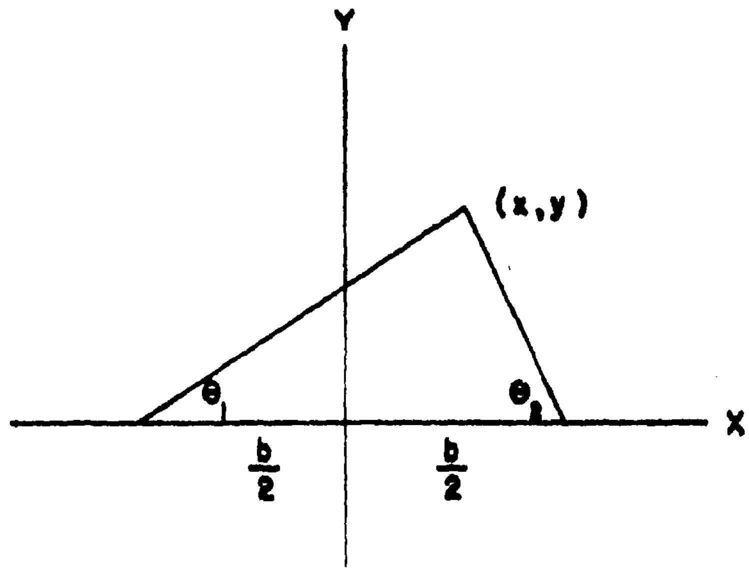


Figure 4a

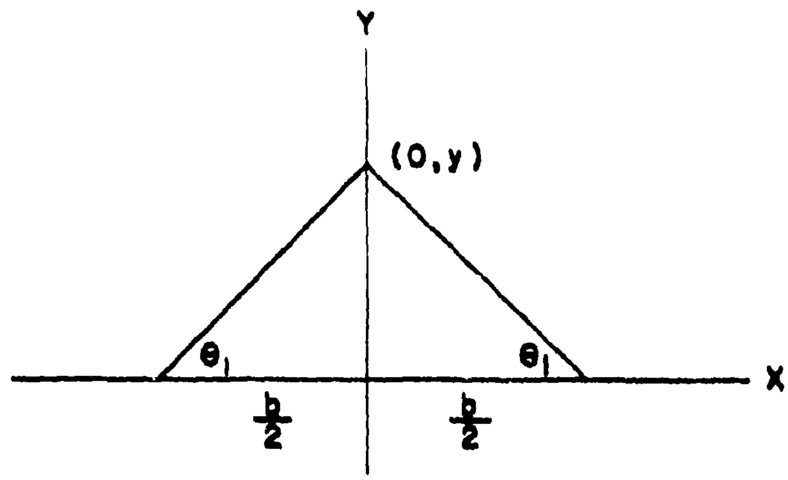


Figure 4b

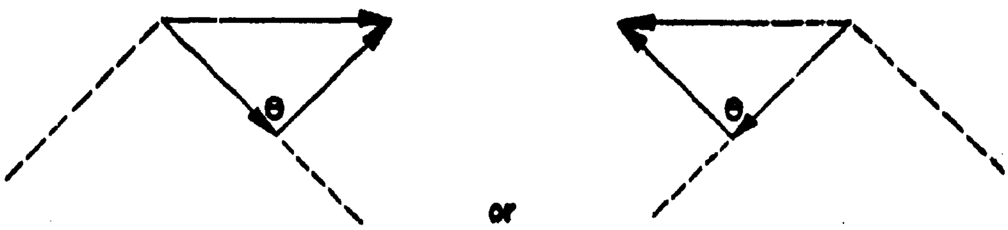


Figure 5a. Same sign

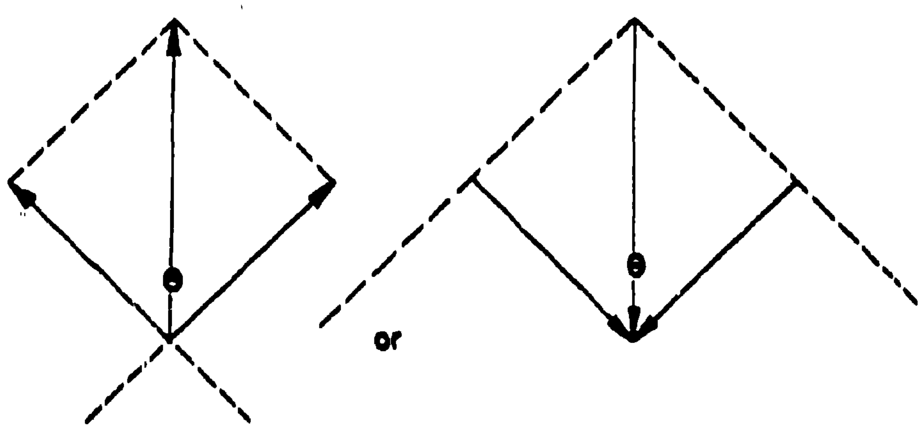


Figure 5b Opposite signs

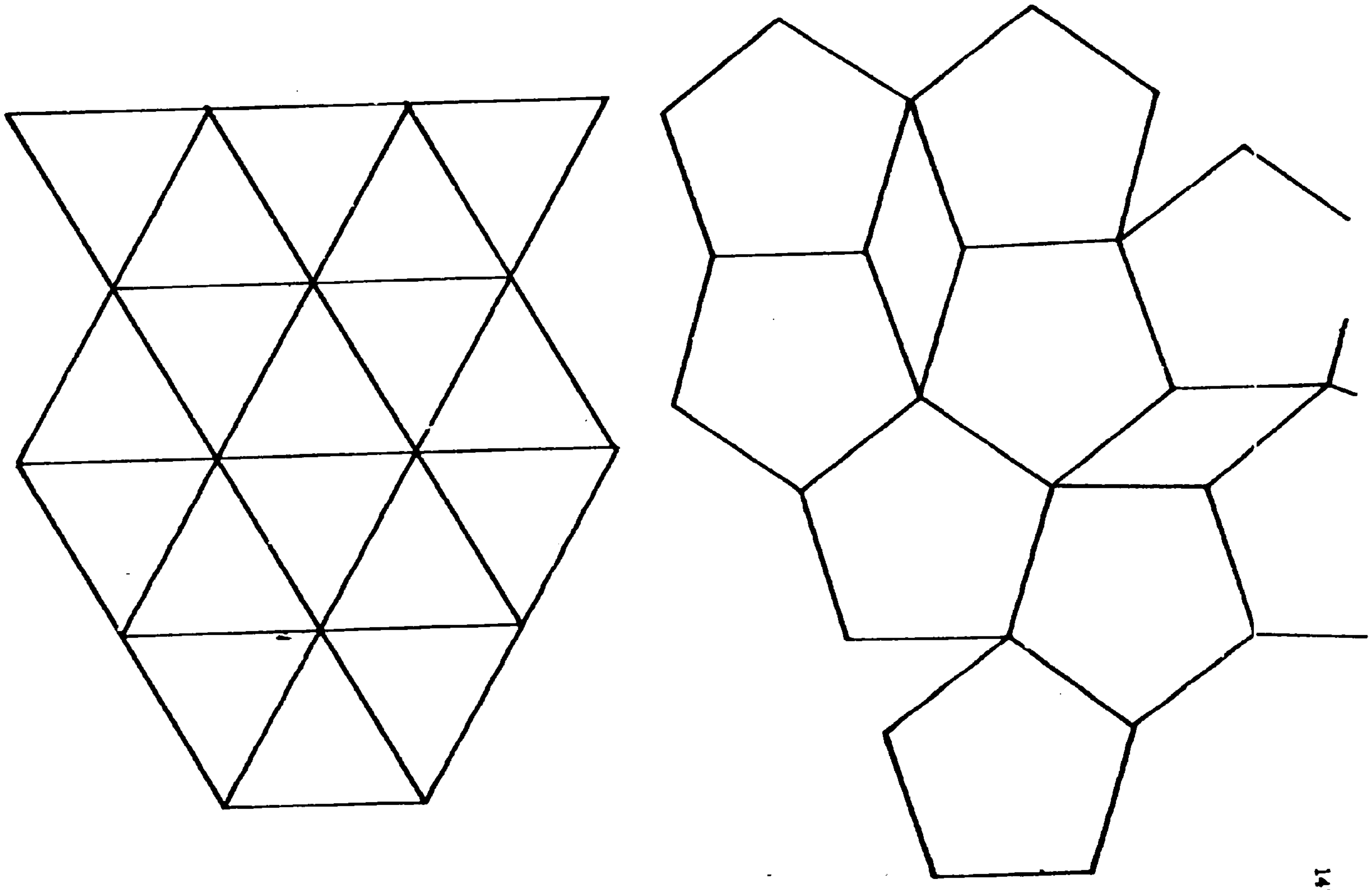


Figure 6 POSSIBLE SYSTEM CONFIGURATIONS

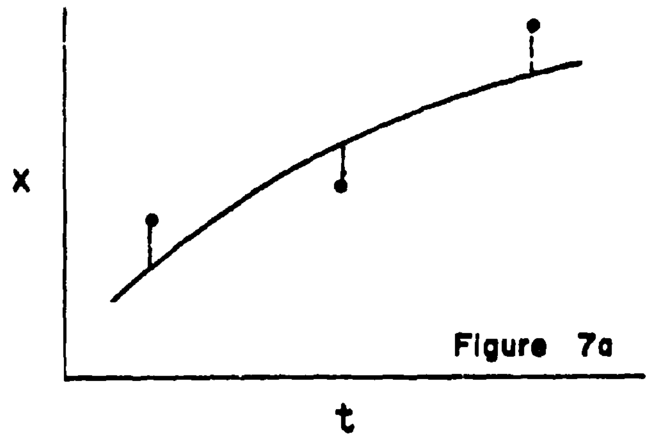


Figure 7a

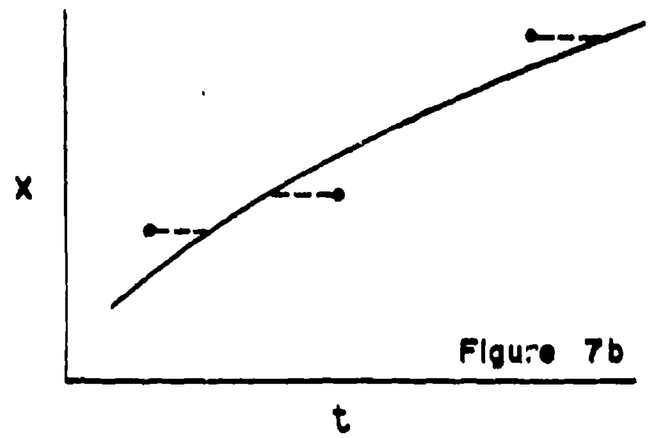


Figure 7b

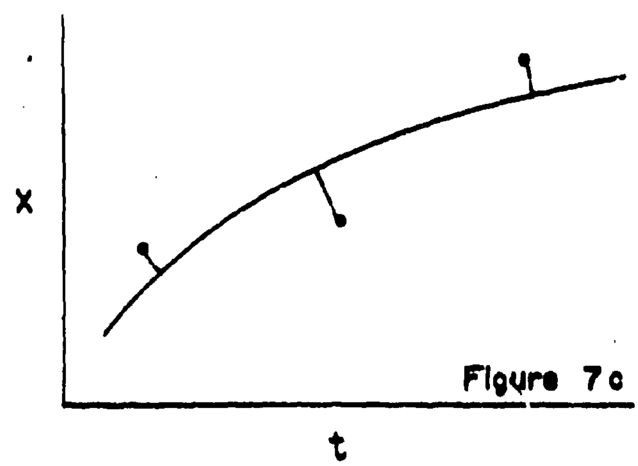


Figure 7c

COMMENTS ON PRESENTATION BY FRED HANSON

Frank E. Grubbs
Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

In my opinion the problems and questions Dr. Hanson raised can be solved satisfactorily only by competent personnel working rather full time on the overall problem! I say this because the problem is so involved from both the physical and the analytical standpoints that it is easy to overlook the importance of all of the "errors" operating simultaneously, so to speak.

Concerning station location geometry, I think that something can indeed be done on this and Dr. Hanson's ideas may be near enough the optimum, considering other involved difficulties. I can see that White Sands might decrease position estimation errors, etc., by optimum station locations, whereas the Atlantic Missile Range cannot really do this.

Just what sums of squares must be minimized, as Dr. Hanson points out, involves considerable study. From my limited experience, I have the feeling that relative time is quite good but that position data is not so good because of intersection geometry, and the errors which creep into this depending on unexplainable biases for the missile flight, calibration, refraction and other corrections, etc. Of course, all of these things vary with the type of instrumentation, etc.

Power spectral density type analyses, are certainly being looked into by many people now and this work is no doubt paying off as many of the problems involved necessarily fall in this area, even though this is an added dimension of complication.

The nearest publication, as Dr. Hanson is aware, which I think is beginning to approach methods required to settle some of the questions Dr. Hanson is raising is the annual report, "Accuracy of AMR Instrumentation", by H. P. Mann. The latest version, as Dr. Hanson knows, does contain a lot of good material and attempts to cover most of the important viewpoints, but still doesn't go far enough.

I think the tracking data analysis problem is by far the most interesting overall one I have been introduced to in recent years, but unfortunately it is something that does not carry the proper priority with many of us in spite of its great importance. Our Panel on Tracking Data Analysis is quite inactive now but if anything comes up on this in the future, I would hope to be in touch with Dr. Hanson.

COMMENTS ON PRESENTATION BY FRED HANSON

Emil H. Jebe
Institute of Science and Technology
The University of Michigan
Ann Arbor, Michigan

Before commenting on Dr. Hanson's two problems, I will first take up the matter of references. I certainly recommend F. S. Acton and K. A. Brownlee (titles Dr. Hanson mentioned). Dr. Hanson has also used Anderson and Bancroft, which is good. Further, I will mention E. J. Williams' "Regression Analysis", J. Wiley & Sons, and Plackett's "Regression Analysis", Oxford Press. Also, O. Kempthorne's "Design and Analysis of Experiments" and H. Scheffe's "Analysis of Variance" may prove useful. There is a book by an Australian, P. G. Guest, "Numerical Methods of Curve Fitting", Cambridge University Press, 1961. Perhaps Dr. Hanson should look at the symposium publication, "Time Series Analysis", SIAM Series in Applied Mathematics, J. Wiley & Sons, 1963.

Now, to Dr. Hanson's problems, Number 1 first. Certainly, I must comment that my experience with the NORC project at Ft. Monroe, 1941-42, and with the Anti-aircraft Artillery Board, Camp Davis, 1942-44, is ancient history by comparison with the state of the art in the 60s. Generally, I agree with Dr. Hanson's analysis of the geometry of the situation, i.e., 45 degree elevation for line-of-sight and nearly orthogonal to missile path for a "reasonable" interval of time. From the algebra associated with the geometry one should be able to work out the error propagation for the position determinations. Of course, one must keep in mind the "best" physical model for the flight path of the missile in using the observed data to obtain best apparent position of missile at a given time.

I tend to think of this first problem more in practical considerations, given that the technical problem of determining location has been resolved to a useful accuracy and precision. Some method of assigning priorities to each day's or each week's missions must be worked out. Then with the resources at hand, an allocation must be made of stations to be manned with selected equipments. Consider Figure 1 for Mission A (highest priority). Enough paired stations, a and a', b and b', etc., must be manned to keep this missile path under adequate surveillance. Now, if a, b, c and d, etc., are too far apart, there will be too much uncertainty in the computed positions in the halfway-between regions. Next, Mission B (second priority) has to be similarly supported at a desired minimum level. If launch

times can be programmed to some extent, it may be that some manned stations can support more than one mission. Continue for say two more Missions C and D. If any resources are left over, consider increasing density of manned pairs for Missions A, B, C and D in that order to shore up obvious weaknesses in trajectory assessment. These practical considerations seem much more relevant to me than going into geometrical considerations beyond the triangle. If is recognized that my sketch implies using rectangles or quadrilaterals in assessing position. When launch times are adequately separated so that all manned stations for each of the four missions can track each launch, then further geometrical considerations may be taken into account along the lines Dr. Hanson has discussed.

Now I turn to the second problem of analysis. Yes, one would like to have variance components for timing error and for position-measuring error. But how can one separate them? Without considerable study, more than I can give at this time, I have no direct suggestion. It is hoped that Dr. Hartley has given Dr. Hanson some useful direct suggestions. I use the term indirect for my ideas because I wish to lean on "design of experiments" considerations. By direct suggestions I mean extracting from present method of collecting data, components of variance of the two kinds desired.

In directing Dr. Hanson's attention to design of experiments concepts, I believe WSMR is in an outstanding position to carry out some special studies. Of course, these activities must be budgeted, but it does not seem unreasonable to program some percent of the WSMR annual budget for R&D on its own job. What the percent should be, I don't know, but 2%, 5% or 7% seems reasonable. Electronics and A/C firms do better. What kinds of experiments one asks? On some missions WSMR may have enough spare resources so that it can double up on position measurements, i.e., re Figure 1, again, put two equipments at each location b, b', c, c', say. I assume that timing errors would be nearly equal at any single location. The smoothed apparent position data (after averaging) should then indicate something about possible "timing component" of error. If a competent person in design of experiments were to spend 3-6 months at WSMR, it seems reasonable that other experiments with useful treatment combinations could be suggested and suitably designed within WSMR's resource frame work.

With respect to the orthogonal regression line, there is nothing in the literature that I am aware of on sampling theory for the regression coefficient or for predicted points. A general reference I recommend is J. B. Coleman, Annals of Math. Stat. 3, 79 (1932). In 1963, I did some work on the design of a flight program carried out in Arizona. By flight replication, we were able to obtain sampling error information about the orthogonal regression coefficient and, thus, overcome the lack of sampling theory based on an internal estimate of error.

Further, as both Prof. Lieberman and I have pointed out, there are no difficulties in obtaining an analysis of variance including a goodness-of-fit term even though the regression fitted is polynomial or otherwise non-linear, so long as the least squares equations are linear in the unknown parameters to be estimated. For the non-linear least squares equations cases, which might arise from a physical model of the missile flight path, I suggest Prof. Hartley's recent paper in Biometrika, 51, 347 (Dec. 1964).

At IST, we have a quite general purpose regression program which is due to Dr. Wyman Richardson. Also, Robert O. Bennett, Jr. and myself are working on a packaged set of sub-routines which can be used for doing Analysis of Variance type calculations. Perhaps, Dr. Hanson should visit us to get information on these programs. Both programs operate on IBM 7090 within University of Michigan Computer Center Executive System.

No doubt WSMR is studying the application and use of the newer high accuracy oscillators for its timing standards. Could not these "atomic clocks" help resolve some of its "timing error" problems? Any WSMR comment on the use of these oscillators will be of interest to us at IST, since we are studying their employment for networks even more widely distributed than those in the WSMR systems.

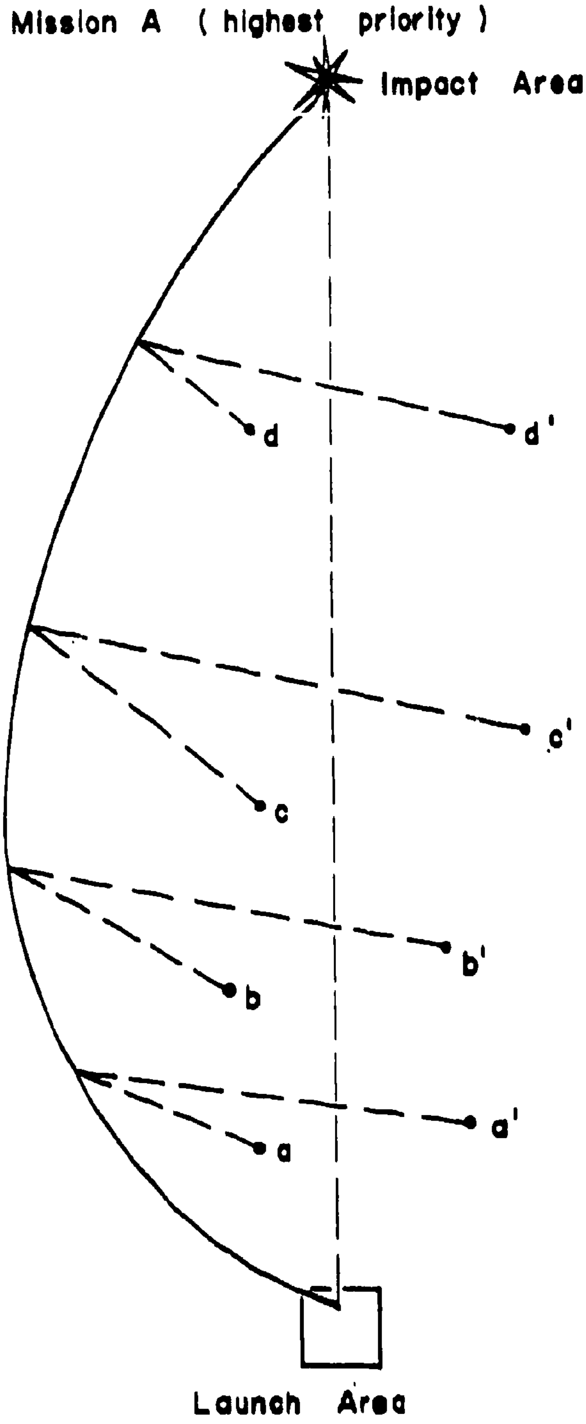


Figure 1
E.H. Jebe

AN EXPERIMENT IN MAKING TECHNICAL DECISIONS USING OPERATIONS RESEARCH AND STATISTICAL METHODS

Andrew H. Jenkins
U. S. Army Missile Command, Directorate of Research and Development
Physical Sciences Laboratory, Redstone Arsenal, Alabama

and

Edwin M. Bartee
University of Alabama in Huntsville,
College of Engineering, Huntsville, Alabama

ABSTRACT. This paper presents a case where decisions are reached and recommendations were made on a multi-disciplined technical research program. The decisions were made on the basis of a technical survey using operations research techniques and statistical methods for evaluation rather than a rigorous technical evaluation of all disciplines. The paper presents the technique used and discusses the practical limitations of the method.

I. INTRODUCTION. The engineer and scientist in government research programs are often required to make decisions and/or recommendations on programs involving advanced technology. Decisions may be required from the individual engineer or a group of engineers. Frequently, the decisions must be made in a minimum of lead time.

The tremendous advances in technology have precipitated a situation where very few research programs are of a single technical discipline. They are usually related either directly or indirectly to other technical disciplines and cannot be treated singularly. A research program, regardless of the number of technical disciplines involved, is an effort to explore and determine the unknown and because of the unknowns is not always conducive to rigorous technical evaluation by an individual or quite often a small group. Certainly, as the number of disciplines increase, the more complex the evaluation becomes.

The engineer, no matter how competent he may be in one discipline, often finds himself making decisions intuitively rather than by rigorous analysis of technical facts. This is so because quite often he does not have the necessary facts, he does not have the time; or he does not have the necessary capability in many disciplines. When the decisions are made intuitively, they are shaded and toned by the engineer's biases, preconceived notions, and past experiences. As the amount of information increases in

a multi-disciplined problem so do his vacillations between biases and preconceptions in the process of making a decision. This condition is accentuated where the research program is such that the technical opinions of others must be considered.

Therefore, what is needed is a systematic approach to the problem, consideration of as many technical factors which may affect the decision as possible, and a method of weighting the factors and quantifying the opinions. In other words, a set of rules are determined and followed systematically until a decision can be reached.

The authors were recently involved in a problem of making a decision and recommendations on certain research programs. The purpose of this paper is to present the approach taken and the use of statistics in the decision making process for an actual case. None of the government agencies or research groups are identified except the U. S. Army Missile Command since the information is for government program planning.

II. BACKGROUND. The U. S. Army Missile Command (USAMICOM) is the technical director of a research program being performed by a research group for the U. S. government. This research program was a multi-disciplined program in missile phenomenology involving theory and experimentation in such disciplines as electromagnetics, optics, plasma diagnostics, microwave-plasma interactions, aerothermochemistry, thermodynamics, fluid dynamics, experimental techniques and instrumentation. This program was one of several similar programs of an overall research program.

The group directed by USAMICOM (identified as Establishment 7) proposed the development and utilization of a larger, much improved hypervelocity launcher of projectiles for research purposes. This among other things precipitated a review of overall research effort in missile phenomenology. In view of this, USAMICOM was requested to give recommendations on the following categorical questions:

I. The past and future utilization of Establishment 7.

II. The need for a large caliber, light gas gun and possible uses in missile phenomenology research.

III. The desirability of building such a gun at some establishment other than 7.

The experimental approach taken by the authors is included except for the coding of all agencies and research groups.

III. THE EXPERIMENT.

A. Design Approach

The purpose of this effort is to provide recommendations in three categories which are of concern to missile phenomenology research programs.

The three categories are as follows:

Category I: The past and future utilization of Establishment 7.

Category II: The need for a large caliber, light-gas gun in missile phenomenology research.

Category III: The desirability of building such a gun at some other establishment.

Due to USAMICOM's close association with past programs and in an effort to carry out this task with minimum bias and maximum objectivity, it was considered appropriate to conduct a technical survey of theoretical and experimental groups associated with such programs. Time limitations permitted only a representative sample of such groups. These groups are known to have knowledge pertinent to all of the above categories.

It was anticipated that a wide variation of data and opinions would be obtained from these groups making orderly, efficient, and unbiased analysis of the survey results difficult. It was decided that a method of analysis based on quantifying of data and opinions must be used. The method selected is the "Case Institute Method of weighting objectives" and is described in Reference 1 in detail.

It was decided to send four engineers as interviewers to visit the selected theoretical and experimental groups. The groups were selected as a representative cross section of those familiar with aeroballistic range techniques and associated research programs, and therefore able to contribute to the resolution of the three categorical problems. The groups were allowed to comment on or off the record to increase responsiveness.

The establishments were visited as shown in Table 1. It can be seen that Interviewer 1 visited Establishments 2, 5, 7, and 11; Interviewer 2 Establishments 3, 4, 9, and 10; Interviewer 3 Establishments 1 and 8; and Interviewer 4 Establishments 6, 12, and 13.

For consistency of the interviews, a master list of questions considered pertinent to the categories was provided to each interviewer and discussed at each establishment. The interviewers recorded a summary of facility data and opinions for use during rating of the factors. Thereby, each interviewer obtained sufficient technical background information upon which he could quantitatively rate ten factors considered pertinent to each category. The ten rating factors for Categories I, II, and III are shown in Tables 2, 3, and 4 respectively. The ten factors were selected as a representative sample which were required to make a systematic evaluation of each category.

The ten factors in Category I were designed to rate Establishment 7 against other establishments. The establishments chosen for 7 to be rated against were 1, 4, 6, 9, 12, and 13. These represented establishments similar to 7 and operated by all government agencies of the Department of Defense, private corporate facilities and an educational institution.

The ten factors in Category III were designed to rate establishments 1, 4, 6, 9, 12, and 13 against 7.

The ten factors in Category II were designed to rate the opinions of both theoretical and/or experimental groups on the need for a large light gas gun.

Each interviewer, after discussion of the factor with the principle investigatory, numerically rated each factor in each category for the establishments visited. These ratings were between 0 and 4. In the selection of a quantitative rating, if the rating was not clearly and easily differentiated from the mean value of 2, the rating was established at that level. This procedure tends to minimize individual bias and enables the survey to approach a truly unbiased conclusion.

B. Factor Rating Criteria

The discussion is confined to the types of information, data and comments obtained for use as a basis for rating the ten factors of each category.

Category I

In Category I the first factor was rated on the basis of the information received on program objectives, types of models required, instruments required, and types of data collected. Also considered was reporting in journals or at symposiums, the opinion of the reporting by other groups, and the degree of success of the program. The rating of the second factor was based on the overall instrumentation capability in flow field visualization, optical radiation, and microwave diagnostic instruments, as well as special instrumentation. The third factor was rated on such criteria as complexity of model shapes, velocity, and data gathering and launching problems. The fourth factor was rated on the basis of type of gun, launch weights, velocities, repeatability, and freedom from malfunction. The fifth factor was rated on the basis of comments of professionals who have had close or personal contact with professionals of Establishment 7. The sixth factor was rated on the basis of the number of available ranges, guns, standard and special instruments, and utilization factor of the facilities. A criteria of minor consideration was estimated capital investment. The seventh factor was rated on a basis of some of the same criteria as factor six plus the ability to initiate programs of widely varying experimental parameters on short notice. The eighth factor was rated on a basis of such things as available space, facility cooperativeness, and facility workloads. Most establishments have existing funded programs planned and limited staff level responsiveness. The ninth factor was rated on relative defense efforts of the establishment. The tenth factor was included on the premise that accomplishments are often proportional to support received.

Category II

In this category an attempt was made during the survey to establish the need for a large caliber gun in missile phenomenology research and to define a large caliber gun. In regard to the large gun proposed reactions varied from "it is feasible" to "it can't be done". Others stated a preference for approaching the possibilities of designing such a gun in small diameter phases, e.g., 2.5 in., 4 in., then perhaps 6 in. It appears from comments obtained that a 3 or 4-inch gun may be the optimum size. A 4-inch gun capable of velocities of 25,000 feet per second would be a size large enough to allow for expansion of the types of experiments which could be performed on an aeroballistic range. A 4-inch gun would also be more easily fabricated, handled, operated, maintained and be capable

of a reasonable firing rate. However, definition of a large caliber gun was a secondary issue, the prime factor being the determination of the real need for a large caliber gun. Factors one and two were rated on the basis of the capibility of a large bore gun to expand the types of experiments and measurements that may be effectively executed under simulated conditions. These factors were most heavily weighted in Category II. The concensus is that this is the foremost justification for a large gun. However, those who expressed this opinion could suggest few programs but some examples are: (1) launching complex geometrical shapes, (2) blast vulnerability studies, and (3) on-board-model telemetry measurements. The fact that new programs cannot currently be suggested does not exclude many suggestions when such a device is available. New types of measurements will be developed in parallel with new types of experiments with larger models. Factor three was only a rating of the opinions of the interviewers on the need for a large bore gun. These opinions vary strongly from favor to disfavor and are reflected in column 3 of Table 6. The X_i column reflects the composite of all factors for each establishment. Factors four and five sought to determine if, in the opinions of others, larger models would improve the thresholds of measurements made by current instruments at a given simulated altitude or provide equal thresholds at a higher simulated altitude. Some respondents indicated that, on a quantitative analysis, significant improvements would not be obtained. Other respondents feel that larger guns would improve thresholds and resolution significantly, especially in optical measurements but not on microwave measurements. Respondents generally agree that simulated data can be more easily utilized in theoretical modeling and computations than in full scale. Some respondents did not feel that this was particularly true to the point of justifying a larger gun than is nominally used, e.g., 1-1/2 inch gun. Some of the respondents to factor seven could not comment, especially if this factor is viewed from the standpoint of a large gun reliability, capital cost, and useful life. Other respondents, even in view of these criteria, feel that more usable data can be obtained at less expense on ballistic ranges than under full scale conditions. The overall response to factors eight and nine varied from neutral on eight to slightly negative on nine. One respondent described quantitatively that examinations of scaling limit increases show that from 10,000 to 20,000 feet of altitude may be obtained by a fivefold increase in size for binary scaling of wake electron densities. Also, only a 20 percent increase in wake lengths that could be scaled would be obtained.

Factor ten was included, at very low weight, merely to emphasize this advantage of ballistic range data gathering when contrasted to full-scale data. While full scale does represent the real case, for purposes of study repeatability is highly desirable. In view of the fact that such diverse opinions and wide variations in responses were obtained, the analysis was made easier by use of the Case Institute Method approach.

Category III

This category assumes that a large caliber gun is needed. It is, therefore, important to determine the best places that such a device should be installed and operated.

The installation of a large caliber gun, which would be heavy, long and cumbersome, would require that the establishment have the necessary heavy moving equipment, transfer locations, and housing to properly operate and maintain it. Factor one considers these present capabilities without new construction.

The installation of a large caliber gun would necessitate increasing the number of persons required to operate and maintain it in a data-gathering program. The operation and maintenance necessitates handling and storage of large amounts of munitions and H_2 of He gas, fabrication of larger models and sabots, telementry packages, and other incidental items required to effectively pursue such a program. In establishments where programs are presently funded to accomplish a mission, such a large program would perhaps overload their present capability. In view of this, the desire of an establishment to participate in a program utilizing a large caliber gun is important. This, in turn, is a function of their interests in the experimental programs to be pursued with a large caliber gun.

The ratio of chamber diameter to model diameter for good compatibility has been estimated between 20 and 30. Therefore, a 5-inch model would require (taking the average) a chamber of 125 inches (approximately 10 feet). Some establishments would require additional chambers for 4-or 5-inch models if this ratio is accepted. Therefore, some establishments may have the desire and interest but not adequate facility and personnel capability or range compatibility.

Other important considerations are the attitude of the establishment to the full - or part-time participation of contractors in data gathering on the range and the participation of contractors intermittently to obtain a few data points of a specific interest. This requires that a certain amount of space on the range for instrumentation be available. Quite often the data can be gathered on shots of opportunity.

In anticipation of research contractor participation, the accessibility of the facility is important to maximum utilization of the facility. In conjunction with this will be the ability to control and direct programs and program changes. Program orientation is also important. It may be desired to pursue a basic long term program with short specific tasks overlaid, the results of which may on occasion change the basic program orientation.

Finally, the cost of a large gun is considered. The overall opinion is that the costs will probably not differ greatly between government establishments. However, an industrial or corporate facility may be more economical than the government facilities.

C. Numerical Analysis

The Case Institute Method of weighting objectives (1) was selected for use in weighting the factors and quantifying the respondent's comments and opinions.

The lack of a universal standard deviation and the small sample dictated the use of Student's 't' distribution for test of significance of the results.

In the Case Institute Method, the ten factors are weighted as follows:

1. One factor in each category is rated most important and given a value of 1.00. Each of the other nine factors are then rated between 0 and 1.00 according to its relatively judged importance.

2. After all factors in a category are rated, the most important factor is compared to the other nine collectively as to importance in the category. If it is judged more important than the other nine collectively, the value of 1.00 first assigned is changed to a value larger than the sum

of the other nine values. If the most important factor is considered to be of the same importance as the other nine, the value for the most important factor should be equal to the sum of the other nine factors. If it is considered to be of less importance than all the other nine, then its value is adjusted to some value less than the sum of the other nine.

3. The most important factor and its weight are established. Next, the second most important factor is compared to the remaining eight. Its weight is established in the same manner as described in 2 above. When the factor's weight is established, the procedure continues to the third, fourth, etc. most important factor until all 10 factors are weighted.

4. This procedure is followed for all three categories.

A composite of the weighting for all categories is shown in Table 8 in order of descending weight. The factors for all three categories can be seen in Tables 2, 3, and 4.

The method of rating the factors was to use the five discrete numerical levels 0, 1, 2, 3, 4. In Category I, each establishment contrasted with 7 was set at level 2 and 7 rated below or above at 0 or 1 and 3 or 4 respectively. In Category II, a neutral position on each factor by the respondent was set at 2 and the degree of disfavor or favor of Category II at 0 or 1 and 3 or 4 respectively. In Category III, 7 was set at 2 for each factor, and each establishment was rated below or above with 0 or 1 and 3 or 4 respectively.

The rating established for each factor in each category was multiplied by the corresponding factor weight and is recorded in Tables 5, 6, and 7. The values are summed for each establishment. In order to normalize the range of response for each establishment in each category, the following equation is used:

$$X_i = \frac{\Sigma(\text{factor wt} \times \text{factor rating}) - 2 \times \Sigma(\text{factor wt})}{2 \times \Sigma(\text{factor wt})}$$

For Category I:

$$X_i = \frac{\Sigma(w_f \times R_f) - 2 \times 3.12}{2 \times 3.12} .$$

For Category II:

$$X_i = \frac{\Sigma(W_f \times R_f) - 2 \times 3.75}{2 \times 3.75}.$$

For Category III:

$$X_i = \frac{\Sigma(W_f \times R_f) - 2 \times 1.49}{2 \times 1.49}.$$

For all categories:

The limits for each X_i in all categories becomes

for $R_f = 0 \quad X_i = -1$

$R_f = 4 \quad X_i = +1$

$R_f = 2 \quad X_i = 0 = \bar{X}'$ (hypothesis value)

$$\bar{X} = \frac{\Sigma X_i}{N}$$

where

X_i = establishment computed response

W_f = factor weight

R_f = factor rating

N = number of establishments.

The sample deviation (S) for each category is

$$S = \left[\frac{\Sigma(X_i - \bar{X})^2}{N - 1}\right]^{1/2}.$$

Student's 't' test for significance is

$$t = \frac{\overline{X} - X'}{S/\sqrt{N}} \quad .$$

Using the data from Tables 5, 6, and 7 for Categories I, II, and III respectively, we calculate the sample standard deviations:

$$S_I = \left[\frac{0.1115}{5}\right]^{\frac{1}{2}} = 0.149$$

$$S_{II} = \left[\frac{1.71}{12}\right]^{\frac{1}{2}} = 0.378$$

$$S_{III} = \left[\frac{.328}{5}\right]^{\frac{1}{2}} = 0.256 \quad .$$

Before the 't' tests are made, a confidence level of 70 percent is set, which is considered appropriate for research (i.e., risk of first kind* $\alpha = .30$) and the following hypotheses are made on each category:

Category I: There is no significant difference in utilization of 7 and other establishments (i.e., $\mu = 0$).

Category II: There is no significant need for a larger caliber gun in the missile phenomenology research program (i.e., $\mu = 0$).

Category III: There is no significant difference between establishments where a large gun should be built (i.e., $\mu = 0$).

The 't' tests are computed for each category:

$$t_I = \frac{.067 - 0}{.149/\sqrt{6}} = 1.105$$

*The risk of rejecting a hypothesis when it is true. Also called the producer's risk.

$$t_{II} = \frac{.148 - 0}{.378/\sqrt{13}} = 1.41$$

$$t_{III} = \frac{.214 - 0}{.256/\sqrt{6}} = 2.05 .$$

The computed values are compared with Student's 't' table values as shown below:

Computed Value for Categories		Degrees of Freedom	Table Value: Percentile Point 70	80	90	95
I	= 1.105	5	0.56	0.92	1.48	2.01
II	= 1.41	12	0.54	0.87	1.36	1.78
III	= 2.05	5	0.56	0.92	1.48	2.01

It can be seen that the tests for all three categories are significant at the original level of confidence of 70 percent which is considered appropriate for advanced research projects. As the tests are significant at this level (the computed value is greater than the table value), all three hypotheses are rejected. The highest level at which the tests are significant and the hypotheses rejected are Category I $\approx$ 80 percent, Category II $\approx$ 90 percent, Category III $\approx$ 95 percent.

If, however, it is considered that the level of confidence should be 95 percent, then the tests for Categories I and II are not significant and the hypotheses accepted. Category III is still significant but inconsequential. For the purposes of decision making in this type research and development programs, the 95-percent level of confidence is considered excessively high by the investigator.

D. Summary and Conclusions

This task is one which is highly complex. Many technical areas of an advanced nature are involved. An honest and sincere effort has been made to reach an unbiased and technically sound solution. The groups queried have provided comments which are spontaneous and which instinctively draw on years of technical experience pertinent to

the problem. Therefore, considerable intellectual attention and technical capability have been concentrated on the three categories. It is not supposed or proposed that every facet has been considered and explored, nor has a rigorously technical approach been used as this would be a formidable task. However, a representative sample of the foremost factors has been considered, and the technical analysis was performed mentally by the respondents.

A systematic approach tothe analysis of a highly complex problem has been used as shown in the numerical analysis. The importance of this approach is the capability to make a decision in the realm of uncertainty and random variation.

Review of the results of the ratings of Category I presented in Table 5 shows that (considering all factors) 7 rates below 9 at -0.178 (or 17.8%) and slightly above all others with 4 and 13 closest with a +0.008 or (0.80%) and +0.024 (or 2.4%), respectively. Comparing 7 to all other establishments for all factors 7 rated at +0.0665 (6.65%) which is significant when compared to the sample standard deviation by the 't' test.

Review of the results of the ratings of all factors for Category II, presented in Table 6, shows that 2 was strongly not in favor of a large gun by a value of 0.701, followed by 10 and 8. Seven was strongly in favor of a large gun with a value of +0.948, followed by 5, 4, 6, and 9. Twelve and 1 were slightly in favor, with values of +0.040 and +0.041, respectively. On an overall comparison of all factors and all establishments there was a favorable response of +0.148 (14.8%). This evaluation does not include the exact launch tube diameters.

Review of the results of rating the factors in Category III, presented in Table 7, shows that 9 with a value of -0.0067 and 13 with a value of -0.0436 compare closest with 7 as the place to build a large gun. Twelve was least favorable with a value of -0.711.

Therefore, on the basis of the analysis of the overall results shown and within the limits of this study the following conclusions were drawn:

Category I

There is an apparent difference in the overall usefulness of 7 compared to other facilities. There is a significantly positive opinion that 7 may be effectively utilized in the future.

Category II

There is an apparent need for a large caliber gun in the missile phenomenology research program. There is a significantly positive opinion that such a device is needed presently and in the future.

Category III

There is an apparent difference between establishments where a large gun could be built and utilized. Establishment 7 is a foremost contender as a desirable establishment for developing the large caliber gun. Recommendations on program continuation together with suggested experiments were made based on these conclusions.

IV. DISCUSSION. The preceding case is a real-world example of how operation research and statistical methods can be utilized to assist in the process of making technical decisions. The particular features of this approach are:

1. An inter-disciplinary team is utilized to bring a variety of technical viewpoints to bear upon the problem.

2. The results of such a team effort are quantified to make it possible for analysis to be made at optimum objectivity.

3. Statistical techniques are applied to evaluate the quantified results.

The key feature of such methods is the concept of risk and probabilistic conditions. Such an approach is particularly useful in the realm of decision-making since the risks are often great and the probabilistic environment is every present. Under such conditions there is no opportunity for drawing a definite conclusion. A decision can only be made at a given level of confidence. The risk of a decision being wrong becomes a calculated part of the problem.

The use of quantitative methods for expressing the results of the experiment can often lead to a process of over interpretation of results often to the neglect of sound technical judgement. Obviously, the decision cannot be made solely with such methods. At best, the decision-maker can be fortified with certain analyses of the experimental results

that will provide a statement of the risk he would take if he should make a decision in one direction or another. Such factual data can often be provided with a minimum of bias from lower echelons so that the decision-maker can benefit from it while exercising his best judgement in the problem.

The experiment was basically concerned with the determination of technical facts that existed within each of the installations. To obtain such facts required us to go through several "bias filters" such as:

1. The ability and willingness of the installation representative (the interviewee) to state the true facts that exist in his group as free of bias and inaccuracies as possible.

2. The ability of the interviewer to gather and transmit the data to the investigator with a minimum of his own personal bias involved.

3. The ability of the investigator to compile the final data as free of his own personal bias as possible.

To accomplish the above purposes is obviously no easy task under any circumstances. The problem was faced in the investigation by utilizing these basic techniques:

1. A multiple of closely related questions were used to conduct the interviewers with each installation representative.

2. The interviewee bias was observed and evaluated by the interviewer in each case.

3. The data was transmitted to the investigator and a concerted attempt was made on the part of the investigator to balance the bias of the interviewer and interviewee through the conduct of an extensive "debriefing" procedure.

4. The bias of the investigator was controlled by both the influence of the interviewer in the debriefing sessions and the systematized method of quantifying the results.

Obviously, the efforts just described could never hope to eliminate all bias and inconsistencies. The recognition of this fact leads us to evaluate the final results with techniques that have been developed for such situations.

We have, in effect, produced quantified results within an enviroment of uncertainty. Such uncertainty is made up of two basic elements. That is, the observed differences in results between installations can be attributed to:

1. Differences that are explained by residual errors and biases that still remain in the experiment in spite of the procedures that were established to eliminate them.

2. Differences that are explained by real effects of the installation on the category in question as far as the study can determine.

The test of hypothesis used in the analysis served to partition these two basic causes of observed differences. To say that a resulting effect was signigicant is to say that, within the limits of this investigation, the observed differences between the selected installations cannot be attributed merely to experimental error. The conclusion is therefore drawn that a real difference exists and a positive conclusion is therefore drawn. It is important to note that for each conclusion there is a comparable level of confidence. Within the realm of an environment of uncertainty all conclusions or decisions must carry this element of risk.

V. REFERENCE

1. Churchman, Ackoff, and Arnoff, "Introduction to Operation Research," John Wiley and Son, New York, New York.

TABLE 1

Establishments Visited by Interviewers

Establishment	Interviewer 1	Interviewer 2	Interviewer 3	Interviewer 4
1			X	
2	X			
3		X		
4		X		
5	X			
6				X
7	X			
8			X	
9		X		
10		X		
11	X			
12				X
13				X

TABLE 2

Establishment Nr. 7 Utilization Evaluation Factors

Category I

Factors are listed in descending order of established weights. Each factor rated with 2 representing each establishment against which Establishment 7 is rated. The rating levels are chosen by this interviewer and the chairman of the survey committee.

1. How do 7's past program results compare to other establishments?
2. How does 7's past instrumentation rate in comparison to other establishments?
3. How did 7's program rate with other ranges in degree of difficulty to perform?
4. How does 7's past gun performance rate in comparison to other ranges?
5. How do 7's professionals compare with professionals of other ranges?
6. How does 7's past facility development rate in comparison to other ranges?
7. How does 7's utility as a data gathering facility in future compare with other ranges?
8. How does future possibility of contractors participation on ranges at 7 compare to other establishments?
9. How strong is 7's desire to continue participation in missile phenomenology research compared with other ranges?
10. How does 7's past funding compare to other range programs?

TABLE 3

Large Bore Gun Evaluation Factors

Category II

Factors are listed in descending order of established weights. Each factor rated at levels between 0, 1, 2, 3, and 4 on basis of data and opinions gathered with 2 representing neutral opinion. The rating levels are chosen by the interviewer and the chairman of the survey committee.

1. Will they expand the types of experiment that may be effectively executed under simulated conditions ?

2. Will they open avenues of new types of measurements ?

3. What is opinion of others doing theoretical work on need for large bore guns ?

4. Will they increase observables levels at higher simulated altitudes significantly ?

5. Will larger bore guns improve reliability and confidence in range measurements ?

6. What is opinion of others on the value of simulated data vs full scale for utilization in theoretical modeling and computations ?

7. How does cost of usable ballistic range data gathering compare with usable full scale data gathering ?

8. Will they contribute significantly to scaling between theory and full scale ?

9. Will they contribute significantly to the establishment of binary scaling limits ?

10. What is the opinion of ballistic range data gathering capability from standpoint of repeatability ?

TABLE 4

Large Bore Gun Location Evaluation Factors

Category III

Factors are listed in descending order of established weights. Each factor rated at levels between 0, 1, 2, 3, and 4 on basis of data and opinions gathered with 2 representing Establishment 7 against which each establishment is evaluated. The rating levels are chosen by the interviewer and the chairman of the survey committee.

1. To what degree are other establishments able to accommodate a large gun from standpoint of housing, operating, and maintenance without facility construction relative to 7?

2. What was capability of other establishments for taking on additional range measurements programs relative to 7?

3. How strongly do other establishments indicate they want to build a large bore gun relative to 7?

4. What was interest of other establishments in taking additional programs relative to 7?

5. To what degree is their present range chamber diameter compatible with large models relative to 7?

6. What is attitude of other establishments toward contractor participation in data gathering on their ranges relative to 7?

7. Is space presently more available on their ranges for contractors' utilization relative to 7?

8. How does accessibility of other establishments compare to 7?

9. How does the ability to control programs at other establishments compare to 7?

10. How will cost of large gun development at other establishments compare to 7?

TABLE 5. CATEGORY I FACTOR RATING

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.58	0.50	0.49	0.40	0.28	0.24	0.20	0.18	0.15	0.10	3.12	
Establishment No.												Sum of Wt x Level	
1	Factor Level (R_f)	2	2	3	3	2	2	2	3	3	2		
	Factor Wt x Level	1.16	1.0	1.47	1.20	0.56	0.72	0.40	0.54	0.45	0.20	7.70	+0.234
4	Factor Level (R_f)	1	2	3	1	2	3	3	2	2	3		
	Factor Wt x Level	0.58	1.00	1.47	0.40	0.56	0.72	0.60	0.36	0.30	0.30	6.29	+0.008
6	Factor Level (R_f)	2	3	2	2	2	3	2	2	2	2		
	Factor Wt x Level	1.16	1.50	0.98	0.80	0.56	0.72	0.40	0.36	0.30	0.20	6.98	+0.119
9	Factor Level (R_f)	1	2	2	0	1	3	2	4	1	3		
	Factor Wt x Level	0.58	1.00	0.98	0	0.28	0.72	0.40	0.72	0.15	0.30	5.13	-0.178
12	Factor Level (R_f)	1	2	3	3	2	4	4	4	1	0		
	Factor Wt x Level	0.58	1.00	1.47	1.20	0.56	0.96	0.80	0.72	0.15	0	7.44	+0.192
13	Factor Level (R_f)	2	2	2	2	2	2	2	2	3	2		
	Factor Wt x Level	1.16	1.00	0.98	0.80	0.56	0.48	0.40	0.36	0.45	0.20	6.39	+0.024
ΣX_i													+0.399
N													6
$\bar{X} = \frac{\Sigma X_i}{N}$													0.0665
$\bar{X}' = 0$ (be design)													0

TABLE 6. CATEGORY II FACTOR RATING

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.92	0.74	0.56	0.40	0.29	0.29	0.18	0.17	0.11	0.09	3.75	
Establishment No.												Sum of Wt x Level	
1	Factor Level (R_f)	3	3	1	0	1	3	2	1	2	4		
	Factor Wt x Level	2.76	2.22	0.56	0	0.29	0.87	0.36	0.17	0.22	0.36	7.81	+0.041
2	Factor Level (R_f)	1	0	0	1	0	1	2	0	0	3		
	Factor Wt x Level	0.92	0	0	0.40	0	0.29	0.36	0	0	0.27	2.24	-0.701
3	Factor Level (R_f)	3	3	1	2	2	1	3	1	3	2		
	Factor Wt x Level	2.76	2.22	0.56	0.80	0.58	0.29	0.54	0.17	0.33	0.18	8.43	+0.124
4	Factor Level (R_f)	3	3	3	1	2	3	3	3	3	3		
	Factor Wt x Level	2.76	2.22	1.68	0.40	0.58	0.87	0.54	0.51	0.33	0.27	10.16	+0.355
5	Factor Level (R_f)	3	3	4	3	4	3	2	3	1	4		
	Factor Wt x Level	2.76	2.22	2.24	1.20	1.16	0.87	0.36	0.51	0.11	0.36	11.79	+0.572
6	Factor Level (R_f)	2	3	3	2	3	3	2	3	2	2		
	Factor Wt x Level	1.84	2.22	1.68	0.80	0.87	0.87	0.36	0.51	0.22	0.18	9.55	+0.273
7	Factor Level (R_f)	4	4	4	4	4	4	4	3	2	4		
	Factor Wt x Level	3.68	2.96	2.24	1.60	1.16	1.16	0.72	0.51	0.22	0.36	14.61	+0.948
8	Factor Level (R_f)	3	1	1	0	0	3	4	1	2	4		
	Factor Wt x Level	2.76	0.74	0.56	0	0	0.87	0.72	0.17	0.22	0.36	6.40	-0.146

TABLE 6. CATEGORY II FACTOR RATING (Concluded)

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.92	0.74	0.56	0.40	0.29	0.29	0.18	0.17	0.11	0.09	3.75	
Establishment No.												Sum of Wt x Level	
9	Factor Level (R_f)	3	3	1	2	3	3	3	1	1	3		
	Factor Wt x Level	2.76	2.22	0.56	0.80	0.87	0.87	0.54	0.17	0.11	0.27	9.17	+0.223
10	Factor Level (R_f)	2	1	1	2	2	1	3	3	3	2		
	Factor Wt x Level	1.84	0.74	0.56	0.80	0.58	0.29	0.54	0.51	0.33	0.18	6.37	-0.151
11	Factor Level (R_f)	2	2	3	2	2	3	4	2	1	4		
	Factor Wt x Level	1.84	1.48	1.68	0.80	0.58	0.87	0.72	0.34	0.11	0.36	8.78	+0.171
12	Factor Level (R_f)	2	2	1	3	2	3	2	3	2	2		
	Factor Wt x Level	1.84	1.48	0.56	1.20	0.58	0.87	0.36	0.51	0.22	0.18	7.80	+0.040
13	Factor Level (R_f)	2	2	3	2	3	3	2	3	2	2		
	Factor Wt x Level	1.84	1.48	1.68	0.80	0.87	0.87	0.36	0.51	0.22	0.18	8.81	+0.175
ΣX_i													+1.924
N													13
$\bar{X} = \frac{\Sigma X_i}{N}$													+0.148
$\bar{X}' = 0$ (by design)													0

TABLE 7. CATEGORY III FACTOR RATING

Factor		1	2	3	4	5	6	7	8	9	10	Sum of Factor Wt	X_i
Factor Weight (W_f)		0.33	0.22	0.21	0.20	0.14	0.12	0.09	0.08	0.06	0.04	1.49	
Establishment No.												Sum of Wt x Level	
1	Factor Level (R_f)	2	1	0	2	2	3	2	3	3	2		
	Factor Wt x Level	0.66	0.22	0	0.40	0.28	0.36	0.18	0.24	0.18	0.08	2.60	-0.128
4	Factor Level (R_f)	2	1	1	1	2	2	2	2	2	2		
	Factor Wt x Level	0.66	0.22	0.21	0.20	0.28	0.24	0.18	0.16	0.12	0.08	2.35	-0.211
6	Factor Level (R_f)	2	2	1	1	2	1	1	3	2	2		
	Factor Wt x Level	0.66	0.44	0.21	0.20	0.28	0.12	0.09	0.24	0.12	0.08	2.44	-0.181
9	Factor Level (R_f)	2	2	3	3	0	1	1	3	1	3		
	Factor Wt x Level	0.66	0.44	0.63	0.60	0	0.12	0.09	0.24	0.06	0.12	2.96	-0.0067
12	Factor Level (R_f)	0	1	0	1	0	0	0	3	2	2		
	Factor Wt x Level	0	0.22	0	0.20	0	0	0	0.24	0.12	0.08	0.86	-0.711
13	Factor Level (R_f)	2	2	1	2	2	2	2	3	2	2		
	Factor Wt x Level	0.66	0.44	0.21	0.40	0.28	0.24	0.18	0.24	0.12	0.08	2.85	-0.0436
ΣX_i													-1.282
N													6
$\overline{X} = \frac{\Sigma X_i}{N}$													-0.214
$\overline{X}' = 0$ (be design)													0

TABLE 8

Factor Weights by Category

Factor No.	1	2	3	4	5	6	7	8	9	10	Σ
Category I	0.58	0.50	0.49	0.40	0.28	0.24	0.20	0.18	0.15	0.10	3.12
Category II	0.92	0.74	0.56	0.40	0.29	0.29	0.18	0.17	0.11	0.09	3.75
Category III	0.33	0.22	0.21	0.20	0.14	0.12	0.09	0.08	0.06	0.04	1.49

IMPROVEMENT CURVES: PRINCIPLES AND PRACTICES

Jerome H. N. Selman, Stevens Institute of Technology,
Rep. the U. S. Army Munitions Command, Dover, N. J.

To build the 1000th B-29 Aircraft took only 3% of the time required to build the first. To build your first window screen or dog house will take you more time than each succeeding one--unless you are a professional window screen or dog-house maker. This feeling is intuitive. The estimation of time reduction for each succeeding item, based upon judgment and experience, is attributed to a human "learning" effect. Mathematically, the way to express this condition would be to use a reduction-type function: A straight line equation with constant negative slope for a constant linear reduction of cost with quantity; a hyperbolic equation with negative exponent for rapid initial reduction of cost with quantity, then slowing down to a limit; more complex equations which are designed to reflect the phases of the specific learning situation.

Models of the cost-quantity relationship, as a predictive technique, came into general use in the airframe industry during World War II after their development in the 1930's. T. P. Wright's pathfinding article* hyperbolically related the average direct man-hour cost to the number of airframes produced. Others have modified Wright's model to show the inverse relationship between the direct labor hours per unit versus quantity produced; this latter formulation being known as the Unit (Improvement) Curve. A linear improvement curve having linear component curves implies that the rate of learning is the same; intuitively, again, the assumption of constant learning rate in all operations is open to question. Wright was of the opinion that different rates of learning are found in the airframe manufacturing process, but he did not inquire into the implications.

Studies in the then-new airframe industry for sub-sonic, reciprocating engine, electrically simple aircraft indicated that although the precentage slope of the improvement curve varied, for every doubling of successive quantities of aircraft, the percentage value was a constant percentage of the unit value of the quantity immediately prior to doubling. The percentage reduction was approximately 80%. This meant that each time the quantity was doubled, the man-hours required to make that designated aircraft was 80% of the man-hours required immediately prior to doubling. Plotting the improvement curve on logarithmic grids gives a "straight line curve",

---

*T. P. Wright, "Factors Affecting the Cost of Airplanes," Journal of the Aeronautical Sciences, Vol. 3, February, 1936, pp. 122-128.

as the grids are so scaled that the interval between doubled quantities are equal; i.e., the distance between one and two is the same as the distance between two and four, or four and eight, or eight and sixteen, etc. Of course, the linear hypothesis should be discarded whenever the unit curves of man-hours and cost depart significantly from linearity--"significant departure" being determined from the slopes of the parallel linear component curves, based on the error permissible in the problem in hand.

Improvement curves are expressed in terms of percentages, such as 80% Curve, 90% Curve, 92% Curve, etc. The percentage figure referring to the fact that man-hours tend to decrease by a definite amount each time the quantity produced is doubled. By correlation and other statistical techniques it has been shown that a graph of the actual performance data (cost, as inferred by man-hours per unit versus quantity produced, or tasks accomplished) may be approximated by a hyperbolic function of the form $y=ax^b$, with a relatively high degree of significance. The fundamental hyperbolic shape is postulated rather than tested (for linearity on double-log scales versus some alternate non-linear functional form for comparison), as a descriptive device for accumulated data. In Improvement Curve terminology, y, is in direct man-hour cost, a, is the direct man-hour cost for "unit Number one", and b defines the "slope" of the curve---"slope" being the ratio of the unit (or average) man-hour cost at two cumulative outputs that differ by a factor of two (2), so that the slope is 2^b. Wright's empirical data on unspecified aircraft yielded a "b"-value of -.322, giving the popular "80% Curve". On arithmetic grid the 80% Curve with a unit one cost of 1000 man-hours is shown in Figure 1, the equation being $y = 1000x^{-.322}$.

To illustrate the mechanics of constructing improvement curves, the 80% Curve will be done in three parts; as shown in Figure 2:

<u>The Unit Time Line</u>: Given a value for any unit P and the slope of the Improvement Curve in percentage form, draw a line from point P through a point X so that it will be twice the unit number of P, i.e., P equals twice X; and the value of X will be the value of P, multiplied by the percent slope of the curve. Equation: $y_i = ax_i^b$ for unit curve.

<u>The Average Time Line Per Cumulative Unit</u>: The Cumulative Average line is drawn in two steps:

1. **The Asymptote.** The Cumulative average line approaches a straight line which is parallel to (after about the 15th unit) and higher than the unit line. To construct the asymptote, obtain the "b" for the improvement curve in question. Draw the asymptote parallel to the unit line so that the values of all points on the asymptote are equal to 1/(1+b) times the values on the unit line. For the 80% Curve, the conversion factor for (1+b) is 0.687, as given on Table I, giving each point on the asymptote a value of 1.475 the corresponding value on the unit line. Equation:

$$\bar{\bar{y}} = \frac{a\,N^{b}}{1+b} \quad .$$

2. The Cumulative Average Line. As an approximation for values between 2 and about 15, the cumulative average values for any unit X is approximately equal to the value shown on the asymptote for, X+3. That is, the average cost of the 4th unit is approximately equal to the value of the asymptote at unit 7. For practical purposes, the average line for units 16 and above may be considered to equal the values of the asymptote. Equation:

$$\bar{\bar{y}} = a \sum_{i=1}^{n} \frac{x_i^{b}}{N} \quad .$$

<u>The Total Line:</u> Draw a line from the value of unit number one to a point at, say, unit number 10, which has a value equal to 10 times the cumulative average value of unit number 10. It is logical that the total time for the first ten units is equal to 10 times the average time (cost) of the first ten units. Equation:

$$Y = a \sum_{i=\frac{1}{2}}^{N+\frac{1}{2}} x_i^{b} \quad ;$$

the corresponding asymptote is N times the cumulative average asymptote, just as the Total line is N times the cumulative average line.

Improvement Curves have been utilized in the Aerospace Industries for Cost estimates, scheduling, efficiency comparisons, procurement and

subcontracting, facilities planning, personnel planning, long-range forecasting, etc., and was proposed for various industries such as home appliances, electronics, construction, machine shops, ship building, etc. The accuracy of the Improvement Curve function as an estimating device is dependent upon a number of factors, including:

Accuracy of Basic Estimate
Choice of the Improvement Rate exponent "b"
Non-linear elements is the real world
Changes in the output rate
Design Changes in product
Influx of "green" manpower
Exit of skilled manpower.

The basic tenet of Improvement Curve philosophy is where there is life (people) there can be learning, the more man-oriented the work, the more learning potential possible. Figure 3 illustrates the generally accepted improvement curve percentages for various man-machines mixes: 75% Man-25% Machine for the 80% Improvement Curve; 50% Man-50% Machine for about 85% Improvement Curves; 25% Man-75% Machine for the 90% Improvement Curves, etc.

Munitions Command Regulation 715-1 requires thorough justification where "program costs are not reduced in accordance with expected learning curve costing." The techniques of the learning or improvement curves can set realistic management goals for setting expected rates of improvement in reducing operating expenses in the Army "Five-Year Cost Reduction Program".

Operations develop trends that are characteristic of themselves. Projecting such established trends is more valid than assuming level performance, or no learning effect. The Improvement Curve function which has remained parochial to the aerospace industries has been presented with the same motive as the rooster who showed his hen an ostrich egg--"It's not that I'm complaining, it's just that I'd like you to see what others are doing!"

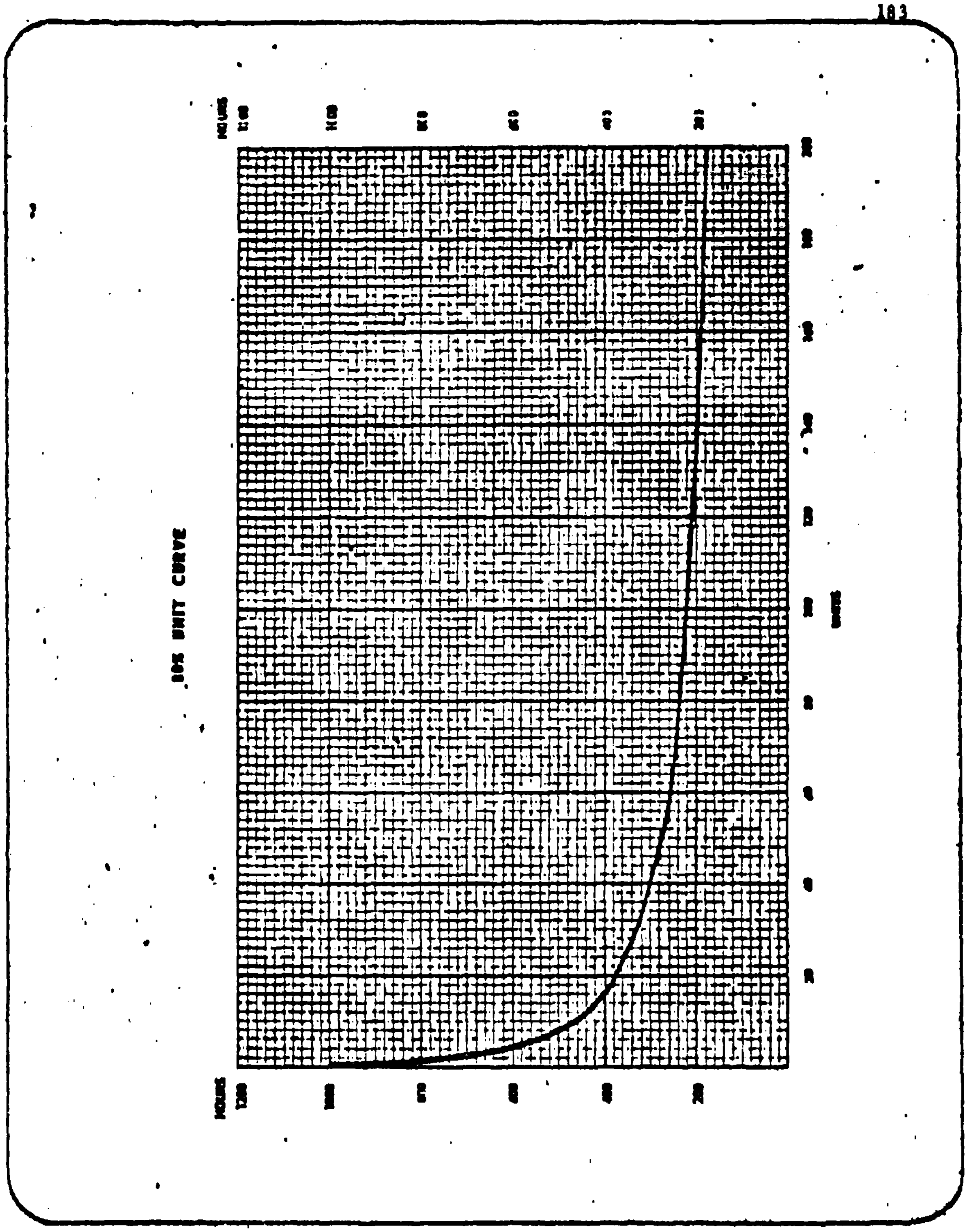


Figure 1

GRAPHICAL CONSTRUCTION OF IMPROVEMENT CURVES
80% IMPROVEMENT

("SLOPE" b = tan^{-1} 17°51' = -0.3219)

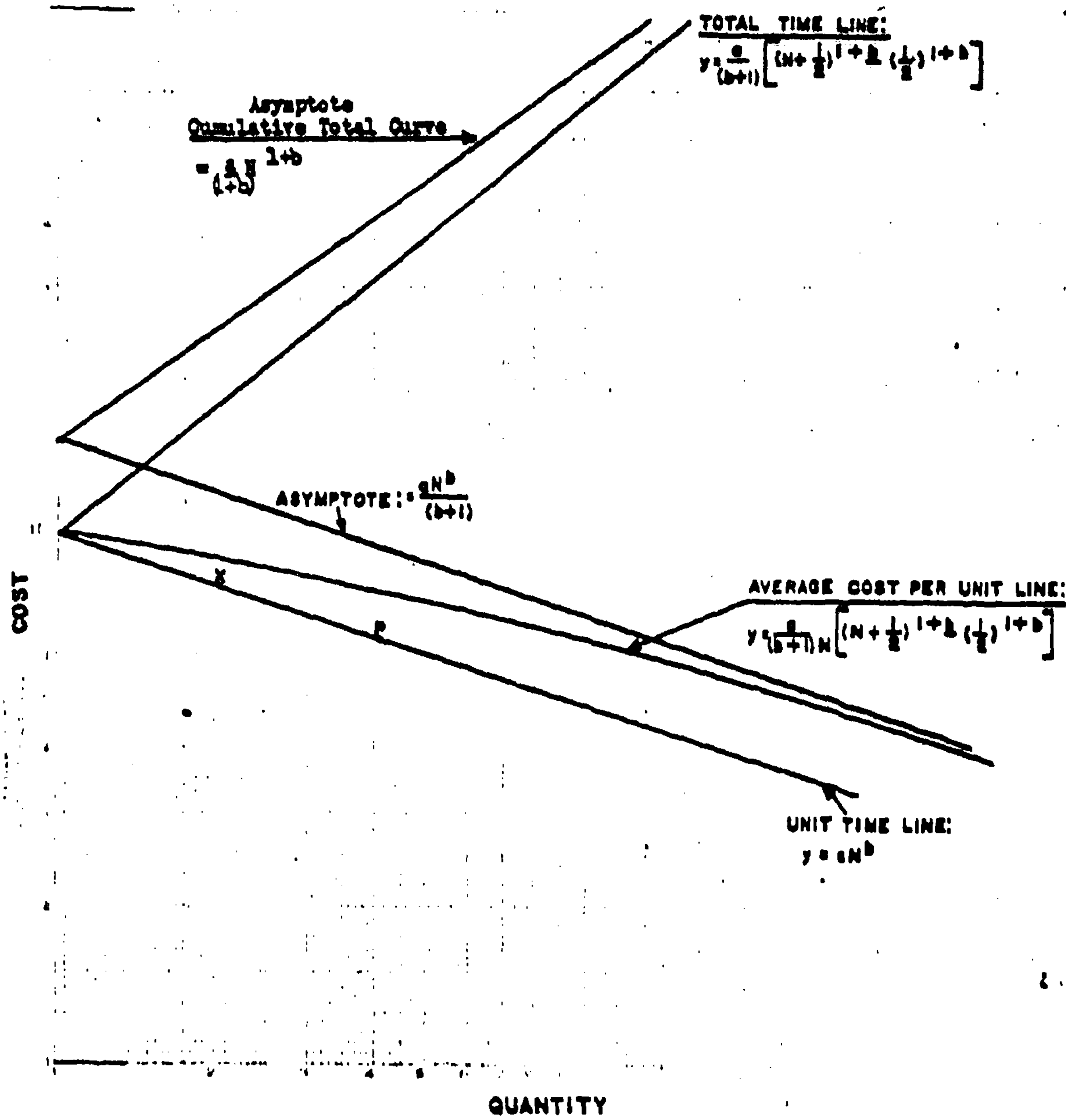


Figure 2

IMPROVEMENT CURVE FACTORS

IMPROVEMENT CURVE, %	b=TANGENT θ	(1+b) CONVERSION FACTOR	θ
50	-1.000	----	45°
55	- .863	.137	40° 47'
60	- .737	.263	36° 24'
65	- .622	.378	31° 52'
70	- .515	.485	27° 14'
75	- .415	.585	22° 32'
80	- .322	.678	17° 51'
81	- .304	.696	
82	- .286	.714	
83	- .269	.731	
84	- .252	.748	
85	- .235	.765	13° 12'
86	- .218	.782	
87	- .201	.799	
88	- .184	.816	
89	- .168	.832	
90	- .152	.848	8° 38'
91	- .136	.864	
92	- .120	.880	
93	- .105	.895	
94	- .089	.911	
95	- .074	.926	4° 14'
99	- .015	.985	0° 50'

Table I

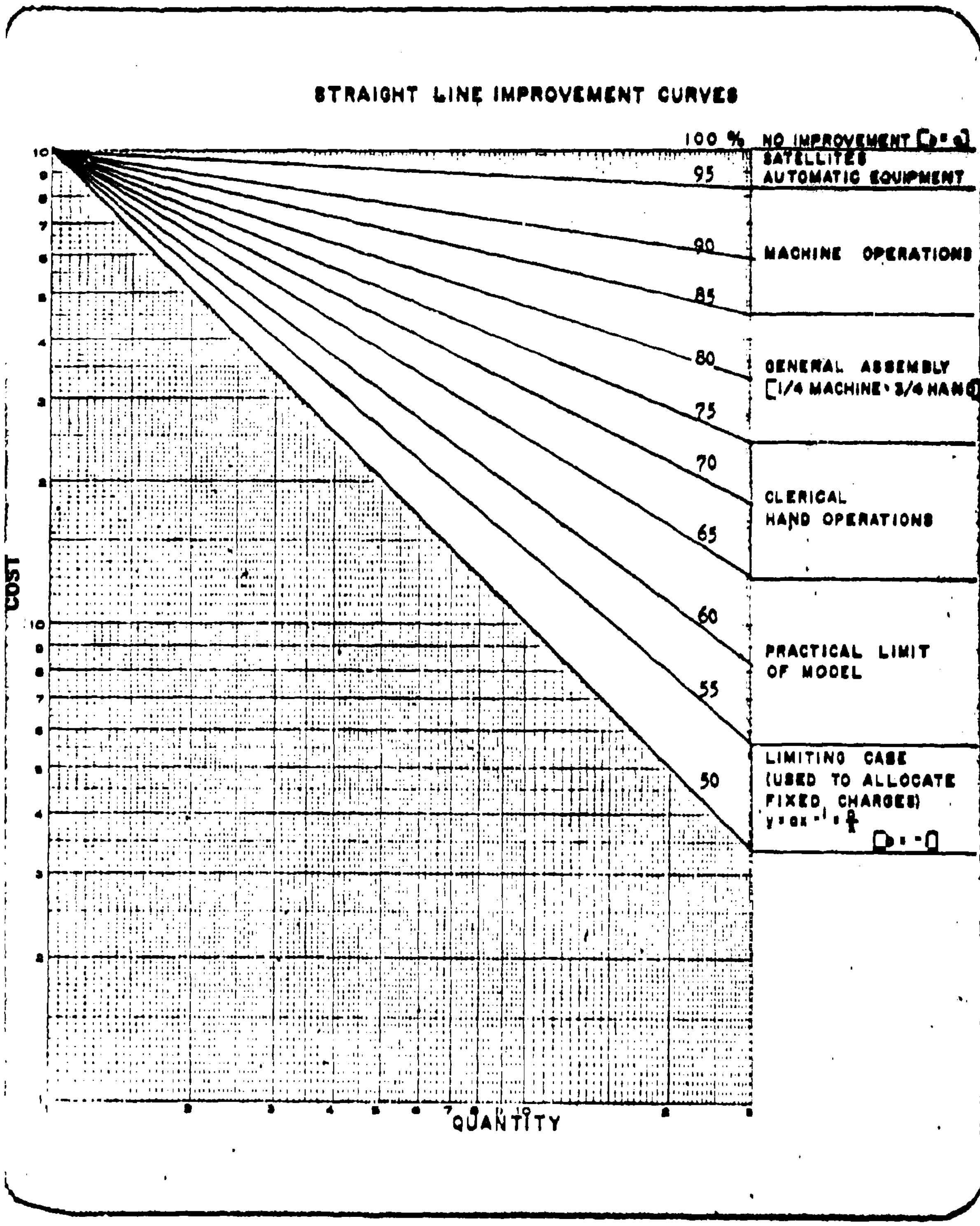

STRAIGHT LINE IMPROVEMENT CURVES
100 %
NO IMPROVEMENT [b=0]
SATELLITES
95
AUTOMATIC EQUIPMENT
90
MACHINE OPERATIONS
85
80
GENERAL ASSEMBLY
[1/4 MACHINE, 3/4 HAND]
75
70
CLERICAL
HAND OPERATIONS
65
60
PRACTICAL LIMIT
OF MODEL
55
50
LIMITING CASE
(USED TO ALLOCATE
FIXED CHARGES)
COST
QUANTITY

Figure 3

THE EFFECT OF RELIABILITY, LENGTH, AND SCORE CONVERSION ON A MEASURE OF PERSONNEL ALLOCATION EFFICIENCY

Richard C. Sorenson and Cecil D. Johnson
U. S. Army Personnel Research Office
Washington, D. C.

Within the United States Army it has been realized for many years that an effective military organization must have the right kind of men as well as the most advanced and effective equipment. Of course this does not mean that the Army must have only the 'best' of the personnel pool, but does mean that those men taken from the personnel pool must be matched with jobs in a way that facilitates maximum manpower utilization. There are two sides to this task of manpower utilization: 1) the various functions performed within the Army must be analyzed to determine the different skills needed to perform those functions, and 2) the individual differences within the personnel pool must be analyzed to find those different abilities that can be reliably measured. At this point we are left with the problem of developing effective measuring instruments and of devising ways and means of assigning men to jobs on the basis of the measure of abilities. This whole attack on manpower utilization rests on the realization that while few men can be trained--no matter how extensive and careful the training--to do all the Army jobs as well as those who do them best, most men accepted by the Army can be trained such that they are effective in performing those skills for which they are most apt, and when properly assigned, will be an asset to the Army.

Thus the solution of the problem rests on successfully accomplishing the following: 1) identifying job families within the Army that require personnel with different ability, 2) identifying and measuring these abilities within the personnel pool, 3) estimating the performance on the job on the basis of measures of ability related to job requirements and 4) assigning men to jobs so as to maximize overall performance.

The first of these steps has been treated in the establishment of the Army occupational areas. Ten occupational areas have been identified and shown to be satisfactory in classifying the various Army functions assigned to enlisted men (EM) [10]*. Recent research indicates that nine

*The numerals in brackets indicate numbered references listed at end of paper.

categories of training schools within Army Advanced Individual training may be differentiated [5]. It may be assumed that continuing research will be required to evaluate the constantly changing functions performed by Army EM as new methods and procedures are introduced.

The Army Classification Battery (ACB) has been developed to measure aptitudes related to Army jobs [4]. An important research mission of USAPRO is to introduce new measuring devices, and to revise and/or validate present tests [7].

The eight current Aptitude Areas are functions of the eleven tests within the ACB and serve as performance estimates for the Military Occupational Specialties (MOS) in one or two occupational categories. These Aptitude Area Scores are currently used for differential classification [10]. (See Figure 1.)

The benefits inherent in differential classification using Aptitude Area Scores stem from the fact that information is obtained relative to the differences in ability between individuals and to differences within the individual. Thus EM may be assigned to jobs for which their probability of success may be a good deal greater than that for Army jobs in general.

The technical gain is twofold. First, a given level of aptitude for a given job can be assured by a lower score on the specific selector highly related to the job than would be required to maintain the same standard of excellence if the selection were based on an instrument less valid for the purpose at hand. Secondly, when recruits are taken above a given cutting score on a general selector, they are removed from that score interval of the aptitude pool for all other jobs as well. However, when recruits are taken above a given cutting score on a specific selector, they come from a much broader range of scores as far as the pool for another specific selector is concerned. To the extent that one specific selector is uncorrelated with a second, the entire range of scores is still available on the latter after selection has been accomplished on the first selector.

Thus we see that for a particular sample of 1800 individuals drawn for the purpose of standardizing a subsequent version of one of the tests 56% were above average on the Armed Forces Qualification Test (AFQT) relative to the original standardization population. Of this same sample, however, 91% were above the average for the Aptitude Area in which they scored highest. (See Figure 2.)

One further operational gain was investigated. Under the former system in which a single test--the Army General Classification Test (AGCT) -- was practically the sole determinant of Army classification, selection for one set of jobs automatically gave those jobs the upper segment of the distribution of test scores. The lower segment was left for the remaining jobs. In the operational problem filling the manpower requirements of an infantry division, approximately one half of the men were combat infantrymen . If, as happened at times during the war, a test were used to select primarily for the noncombat specialties, these jobs would be filled by using the upper half of the distribution . In such a case only the lower half of the distributions of test scores would be available for the combat jobs as indicated in Figure 3. However, when the results of the distribution of men into aptitude areas corresponding to job families for the infantry division used in the standardization study mentioned above are viewed, the distribution of AGCT scores for the non-priority or combat jobs is seen to be almost equal to the distribution of AGCT scores for the priority jobs. This is shown graphically in Figure 4.

A great deal of research has been undertaken to make optimal allocation feasible. Various versions of the optimal regions and other methods are now available for operational use [1] . In the research reported in this paper a routine derived from the Hungarian solution to the transportation problem was used [8] .

In this paper we will be concerned with investigating characteristics of performance estimates (and the test battery from which they were derived) as they relate to the criterion of personnel allocation efficiency as measured by the average performance under conditions of optimal allocation. This measure of performance is the objective function to be maximized in the transportation problem. Many relationships involving this objective function and the variables of this study may easily be calculated analytically assuming ideal conditions, e.g., continuous normally distributed psychological test scores. For instance Brogden [2, 3] has shown that when other factors are held constant and certain conditions assumed, the efficiency of allocation is directly proportional to the validity of the performance estimate, and that one may determine by analytic means the allocation efficiency for given numbers of jobs, percent of personnel pool rejected, and intercorrelation of performance estimates. In reality, however, we are not dealing with continuous variables and frequently other assumptions are not met. Also, in practice the scores are often

transformed in such a way that considerable information is lost. It is less easy to investigate the more realistic situations analytically. Thus we have embarked on a program to study by a Monte Carlo approach the general relationship between amount of information in a distribution of discrete performance estimates and the performance level it is possible to achieve by the most efficient pattern of personnel assignments.

The basic step in the implementation of a statistical experiment is the generation of uniformly distributed random numbers. We have used computer routines which generate pseudo-random numbers by the power residue method [9]. These distributions of uniform variables are then transformed to distributions of normal variables. This transformation results in a matrix, X, of order n by k, i.e., n entities are represented each by a vector of k simulated scores:

$$(1) \qquad X = \begin{bmatrix} X_{11}, X_{12}, \ldots X_{1k} \\ X_{21}, X_{22}, \ldots X_{2k} \\ \vdots \\ X_{n1}, X_{n2}, \ldots X_{nk} \end{bmatrix}$$

where

$$(2) \qquad \left. \begin{array}{lll} & X'X & \rightarrow nI \\ \text{and} & 1'X & \rightarrow 0 \\ \text{when} & n & \rightarrow \infty. \end{array} \right\}$$

We see then that for each sample we generate a matrix that has an expectation for its covariance matrix of the identity matrix.

Now we desire to further transform the matrix X by post multiplication by a matrix T such that the resulting matrix has for its expected covariance matrix a given matrix C:

$$(3) \qquad \left. \begin{array}{lll} & XT & = Y \\ \text{where} & Y'Y & \rightarrow nC \\ \text{when} & n & \rightarrow \infty \end{array} \right\}$$

The matrix C is specified as a function of the desired standard deviation and intercorrelation of the variables:

(4) $$C = s R s,$$

Where R is the desired correlation matrix and s is the diagonal matrix of standard deviations.

We wish to find the matrix T such that the conditions in (3) will hold. From these equations we may write the requirement that:

(5) $$\left(\frac{1}{n}\right) Y'Y = \left(\frac{1}{n}\right) T'X'XT \rightarrow C$$

when $n \rightarrow \infty$.

From (2) we see that

(6) $$\frac{1}{n} X'X \rightarrow I$$

when $n \rightarrow \infty$

and from (5) and (6) we have

(7) $$T'T = C\,.$$

We may represent the matrix C in terms of its basic structure:

(8) $$C = Q\Delta Q'$$

where $QQ' = Q'Q = I$.

We know that the matrix C to any power e.g., ℓ, may be formed by raising the eigen values of C to that power, premultiplying by Q and post-multiplying by Q′[6]:

(9) $$C^{\ell} = Q\Delta^{\ell}Q'$$

(10) thus $$C^{\frac{1}{2}} = Q\Delta^{\frac{1}{2}}Q'.$$

Formula (10) could be demonstrated as follows:

$$C^{\frac{1}{2}} C^{\frac{1}{2}} = Q\Delta^{\frac{1}{2}} Q' Q\Delta^{\frac{1}{2}} Q' = Q\Delta Q' = C. \tag{11}$$

We will let

$$T = C^{\frac{1}{2}} \tag{12}$$

We see that

$$T'T = C^{\frac{1}{2}} C^{\frac{1}{2}} = C \tag{13}$$

Hence a transformation solved for by equation (11) meets the requirement of (7) and while there are an infinite number of transformations that meet this requirement the one indicated is by far the most advantageous since it provides for uniformity of rounding errors and impartially improves normality of the transformed scores.

Thus we may simulate samples of personnel by building into the score distribution characteristics of performance estimates in which we are specifically interested. These performance estimates may in turn be a function of such test characteristics as length, reliability and validity. The effectiveness of a test or of the resulting performance estimation is determined by its potential contribution to the optimal allocation average, that is, the average estimated performance of men on the jobs to which they are assigned.

Let us first consider one of these characteristics of a distribution of performance estimates: namely the standard deviation. Often times, in the course of personnel operations where men are actually being assigned to jobs on the basis of measured attributes, distributions of scores are transformed from distributions in which there are two or three significant digits to distributions in which there is only one significant digit. This is the case in assigning men to jobs in the Army. The three digit Army Aptitude Area Score is coded according to AR 611-259 to a score taking on the values ranging from zero to nine. The questions we ask are: 1) What loss of information occurs when scores are coded to a one digit scale, and 2) What affect does this loss of information have on average performance when these scores are used to assign men to jobs?

In Figure 5 we demonstrate the effect of coding the scores of a continuous distribution centered at 50 into nine score scales, e.g., entities with scores less than 11.5 were given a coded score of 1, entities with scores 11.5 or greater but less than 22.5 were given a coded score of 2, ... entities with scores 88.5 or greater when given a score of 9. The upper portion is the resulting distribution when the original distribution has a standard deviation of 20. The information measure, H, has an intuitive appeal because it is sensitive to both the size of the coded interval and the spread of scores. For the above distribution H may be calculated by

$$H = \sum_{i=1}^{9} (p_i \log p_i) \tag{14}$$

where p_i is the proportion of the entities in the ith interval and $\log p_i$ is the natural logarithm of p_i. The information measure corresponding to the distribution represented in the top of Figure 5 is 1.991. In the lower figure, a similar transformation was performed on a continuous distribution, where the original distribution has a standard deviation of 10. We see here that the cases are primarily distributed in intervals 4, 5, and 6, that they are much more closely grouped together. That much more information is lost is indicated by the corresponding information measures which is 1.372. We may note that the maximum value for the information measure corresponding to a nine score scale is 2.197 which occurs when the distribution is uniform.

Now we can easily see that information is lost when we go from several significant digits to one significant digit. We also see that more information is lost when the standard deviation of the parent distribution is small than when it is large. We desire to investigate the degree to which such information loss affects the optimal allocation average.

Another variable of interest is the quota restriction places on the optimal allocation. A natural quota is defined as the number of men that would be assigned to a job if everyone were assigned so as to maximize his individual performance without regard to quotas. In the case of equal variances and intercorrelations among performance estimates, the natural quotas are equal, i.e., uniform. On theoretical grounds we can conclude that the degree to which the quotas are perturbed from the natural is related to the allocation average. However, the effect of this quota factor

on the other relationships must be studied empirically. We see in Figure 6 the percentage quotas imposed on optimal allocation for the situation where we have 16 jobs and where we simulate only 4 jobs. Note that the natural or uniform quota for 16 jobs is .0625. That is the proportion of the total personnel pool that would be allocated to each job. For 4 variables it is .25. There are two considerations that determined the perturbed quotas. The first was that we wanted at least one individual to be assigned to each job, for both the 16 and 4 variables for each of the sizes of samples. The second was that we wanted the ratio of the information measure that was found to exist between the 4 and 16 variable situation, for the natural quotas, to exist also for the perturbed quotas. We required that the uncertainty of assigning men to jobs with 16 variables be twice that for assignment with 4 variables for both the natural and perturbed quotas. The resulting proportions indicated in the table were the result of the two considerations mentioned above. We feel that in imposing these quota restrictions in our experiment we are being realistic, in that the necessary perturbations in the quotas in the actual operational conduct of the Army personnel system would not be greater than this.

In order to study these effects, a 2^5 factorial experiment using simulated performance estimates was designed. The five factors were: (1) standard deviation of the estimated performance; (2) number of cases in the sample; (3) number of variables; (4) number of score intervals; and (5) quota restriction. Figure 7 indicates the various levels of the five factors that were used. The performance estimate variables were generated such that they had an expectation of .70 for their intercorrelation. For those samples that were randomly assigned to Level a of Factor 1, the parent distribution was generated to have a standard deviation of 10; for those assigned to Level b, the standard deviation was 20. Similarly, those samples assigned to the first level of Factor 4 were transformed to have 9 score intervals, while those assigned to Level b were transformed to have 99 score intervals. The number of cases and variables represented correspond to the level of Factors 2 and 3 to which the sample was assigned. Those samples assigned to Level a of Factor 5 were allocated with uniform quotas. Those samples assigned to Level b were allocated with perturbed quotas. Thus we have a 2^5 factorial experiment in which there are 32 cells. The experiment was initially replicated 10 times. Three hundred and twenty samples were generated from a simulated personnel pool and allocated optimally to either 4 or 16 job categories. Figure 8 is a flow diagram indicating the five steps in this experiment. In step 1, the matrix X of normally distributed random numbers, was generated. In the second

step, the matrix Y of continuous performance estimates, was derived by multiplying the matrix X by the transformation matrix. The continuous performance estimates were used in evaluating the allocations under the various experimental conditions by averaging the estimated performance of men on the jobs to which they were assigned. In doing this, we used the continuous performance estimates, since continuous performance estimates yield an unbiased estimate of the actual performance of men on the job, whereas discrete performance estimates would have introduced a slight bias. As may be seen from the arrow going from step 2 to step 5 in the graphical presentation, the continuous performance estimates were used in the calculation of the allocation average. In step 3, the matrix $\tilde{Y}$ was derived by forming a discrete performance estimate from the continuous performance estimate. This was done simply by forming the scores into either 9 or 99 score intervals. Step 4, the allocation step, was accomplished by a computer program which optimally allocates men to jobs by a linear program derived from the Hungarian Solution to the transportation problem [8]. The average performance for men who are thus allocated is then calculated. It is these allocation averages which are subjected to the analysis of variance in this experiment.

We have put the analysis of variance to a slightly different use in our experiment than is the usual case. Theoretical considerations in this experiment dictate that we should expect significant differences between the two levels of each of these five factors. We are not testing to see if the null hypothesis should be rejected, but we are performing the analysis of variance so that in the event that the main effects are not significant, we can evaluate our simulation for its adequacy with regard to the number of replications. Thus, the purpose of the analysis of variance in this experiment is primarily that of evaluating the number of replications that we used in our simulation. With 10 replications, four of the five factors were highly significant at the .001 level or less. However, the effect of Factor 2, the number of cases in each simulated personnel sample, was not significant. We then repeated the experiment using as the level of Factor 2 different sizes of samples: 32 and 192. We found that while there was a small difference, this difference was insignificant both statistically and practically. We conclude that when allocating large quantities of men to jobs under the conditions specified above, we are justified in sub-optimizing (random sampling the overall sample into several subsamples and allocating each of the subsamples optimally). In so doing, we may operate with less computer space with little concern for the loss in allocation average.

In Figure 9 we have shown the mean performance for the levels of those factors that were found to be statistically significant. The results indicated that the number of variables is the most important of the factors of the experiment. We could increase the gain over random allocation by 72% by increasing the number of criterion variables from 4 to 16. This indicates that one of the most promising avenues of psychometric and personnel research is to differentially predict more job categories or job families than we are now doing. The number of score intervals factor was a significant one as was the quota factor. However, the latter was of no practical significance. We feel that we may continue to use natural (or uniform) quotas in our research work and generalize our interpretation of results to realistic situations where the quotas are not uniform.

The interactions of Factor 1 with Factor 4, and Factor 3 with Factor 4 were both significant at the .01 level. The cell means for these two interactions are found in Figures 10 and 11. It appears that the information loss is considerably more crucial when we are dealing with 16 differential job predictions than when we are dealing with only 4. The significant interaction between Factor 1 and Factor 4 indicates that the loss in the allocation average going from 99 score intervals to 9 score intervals is much greater when the standard deviation is 10 than when it is 20. (Recall that this was predicted from considerations of the amount of information in the respective distributions.) The results thus far indicate that: (1) mean performance may be increased by increasing the number of differential performance estimates, (2) when attempting to do # 1, it is important that all the information possible be retained in the score distribution by using as many score intervals as is meaningful, and (3) in going from a 99 interval distribution to a 9 interval one, the loss is doubled if the original standard deviation is 10 rather than 20.

These results may be evaluated from at least two points of view: first, from that of an agency dealing with actual score distributions, and second, from the point of view of the test constructor. He looks at our number of intervals factor as the number of items in a test, since the number of meaningful score intervals is related to the number of test items. Furthermore, he may consider our standard deviation factor in terms of the relationship between the standard deviation of a test and the reliability and number of items in the test.

Upon consideration of the factors mentioned above, an additional experiment was designed. The factors to be studied and their levels are indicated in Figure 12. Ten samples of 200 entities were assigned to each of the eight cells of the design formed by the first three factors. Each sample was optimally allocated and evaluated at each level of Factor 4. For each sample, vectors of test scores were generated and transformed to represent perfectly valid performance estimates.

Figure 13 represents by a flow diagram the steps followed in the experiment. First, the matrix of normal random numbers, X, was generated. In step 2, X was transformed to a matrix of continuous test variables. In step 3 the continuous test variables were formed which were to be used in the evaluation of our allocations in step 8. In step 4, the discrete test variables, $\tilde{G}$, were formed from the continuous test variables, matrix G, by creating either 20 or 40 discrete score intervals. From $\tilde{G}$, the performance estimates, $\tilde{Y}$, were formed by the appropriate regression equation. These performance estimates were used in allocating the men to jobs in step 7. In step 6, the performance estimates were transformed to stanine form and again the men were allocated to jobs and the allocation was evaluated.

Note that this analysis of variance is a split plot analysis of variance in which we can analyze the between-samples variance and the within-samples variance. First, let us look at Figure 14, which reports the results of the between-samples variance. The effect of intercorrelations, reliability, and the inter-action between intercorrelations and reliability, were all significant. The number of items was significant only at the .25 level, with 10 replications. We see from the analysis of the within-samples variance (see Figure 15) that the score conversion factor was significant and the score conversion-reliability interaction was significant as were the three factor interactions of score conversion, intercorrelation, reliability and score conversion, reliability, number of items. Let us now look at the difference in the mean job performance for the two levels of each of the four factors as indicated in Figure 16. It is of interest to note that by reducing the intercorrelation among the test variables, a great increase can be brought about in the allocation average (i.e., mean job performance). We see also, that the test reliability is an important consideration. Let us note that the difference in mean performance for the two different levels of number of items, apart from validity, inter-correlation, and reliability, was in the direction that the larger the number of items, the higher the allocation average. The difference across the two

levels of score conversion (i.e., no conversion vs. a conversion from the score to the stanine) was also a significant one. As we look at the interaction between the intercorrelation among the test variables and the reliability (see Figure 17), we see that the reliability is a more crucial consideration when high intercorrelations prevail than when they are low.

Inasmuch as we did not find the number of items to be a significant consideration, we replicated the experiment for crucial cells 20 more times. In Figure 18, we see the results of that analysis of variance. We see that the number of items is significant, and that the score conversion as well is statistically significant. In looking at the means for that experiment, we find that as we go from 40 items to 20 items, that is, when we cut the length of the test in half, even if we would keep the reliability of the test the same and the validity of the test the same, we would lose approximately 8% of our gain over random allocation of men to jobs.

The results of this work indicate that the use of caution is warranted in advocating the use of shorter tests in optimal differential classification, even if the shorter tests retain the reliability and validity of the longer tests, especially if the reliability of the tests is closer to .7 than to .9. This and other research currently in progress has impact on the planning of further test development research and on the operational handling of test scores and performance estimates. Furthermore, it demonstrates that simulated experiments can yield information concerning possible trade-off between allocation average, testing costs, and the relative costs of test development. Even more efficient experiments could be done to estimate the magnitude of differences by employing variance reduction methods. One, the regeneration of the same sample transformed for each cell in the design, would be especially appropriate for this type of study. It was not used in this project because the model for analysis of variance does not provide for a residual estimate of variance. Future projects will employ variance reduction techniques.

REFERENCES

1. Boldt, Robert F. Development of an Optimum Computerized Allocation System. Technical Research Report 1135, U. S. Army Personnel Research Office, OCRD, DA, 1964.

2. Brogden, Hubert E. Increased Efficiency of Selection Resulting from Replacement of a Single Predictor with Several Differential Predictors. Educ. and Psychol. Measurement, Vol. 11, No. 2, 1951.

3. Brogden, Hubert E. Efficiency of Classification as a Function of Number of Jobs, Percent Rejected, and the Validity and Intercorrelation of Job Performance Estimates. Educ. and Psychol. Measurement, Vol. 19, No. 2, 1959.

4. Helme, William H. Differential Validity of the ACB for Courses in Seven Job Areas. PRB Technical Research Report 1118, Personnel Research Branch, R&DD, TAGO, 1960.

5. Helme, William H., and Fitch, D. J. Grouping Army Training Courses by Army Classification Battery Factors. Technical Research Note 128, U. S. Army Personnel Research Office, OCRD, DA, 1962.

6. Horst, Paul. Matrix Algebra. Holt, Rinehart and Winston, Inc., New York, 1963 (see Section 17.7).

7. Johnson, C. D., Waters, L. K., Helme, W. H. Factor Analysis of Experimental Noncognitive Measures of Combat Potential. Technical Research Note 147, U. S. Army Personnel Research Office, OCRD, DA, 1964.

8. Kuhn, H. W. The Hungarian Method for the Assignment Problem. Naval Research Logistics Quarterly, Vol. 2, Nos. 1 and 2, 1955.

9. Shreider, Yu A. Statistical Testing. Elsevier Publishing Co., New York, 1964 (p. 199).

10. Zeidner, Joseph, Harper, B. P., and Karcher, E. K. Reconstruction of Aptitude Areas. PRB Technical Research Report 1095, Personnel Research Branch, PR&PD, TAGO, 1956.

ARMY CLASSIFICATION BATTERY

TEST	SYMBOL
VERBAL	VE
ARITHMETIC REASONING	AR
PATTERN ANALYSIS	PA
CLASSIFICATION INVENTORY	CI
MECHANICAL APTITUDE	MA
ARMY CLERICAL SPEED	ACS
ARMY RADIO CODE	ARC
GENERAL INFORMATION	GIT
SHOP MECHANICS	SM
AUTOMOTIVE INFORMATION	AI
ELECTRONIC INFORMATION	EI

ARMY APTITUDE AREAS

TITLE	SYMBOL	FORMULA
INFANTRY - COMBAT	IN	$\frac{AR + 2CI}{3}$
ARMOR, ARTILLERY, ENGINEERS-COMBAT	AE	$\frac{GIT + AI}{2}$
ELECTRONICS	EL	$\frac{MA + 2EI}{3}$
GENERAL MAINTENANCE	GM	$\frac{PA + 2SM}{3}$
MOTOR MAINTENANCE	MM	$\frac{MA + 2AI}{3}$
CLERICAL	CL	$\frac{VE + ACS}{2}$
GENERAL TECHNICAL	GT	$\frac{VE + AR}{2}$
RADIO CODE	RC	$\frac{VE + ARC}{2}$

Figure 1. Army Classification Battery (ACB) tests and Army Aptitude Areas as functions of ACB variables.

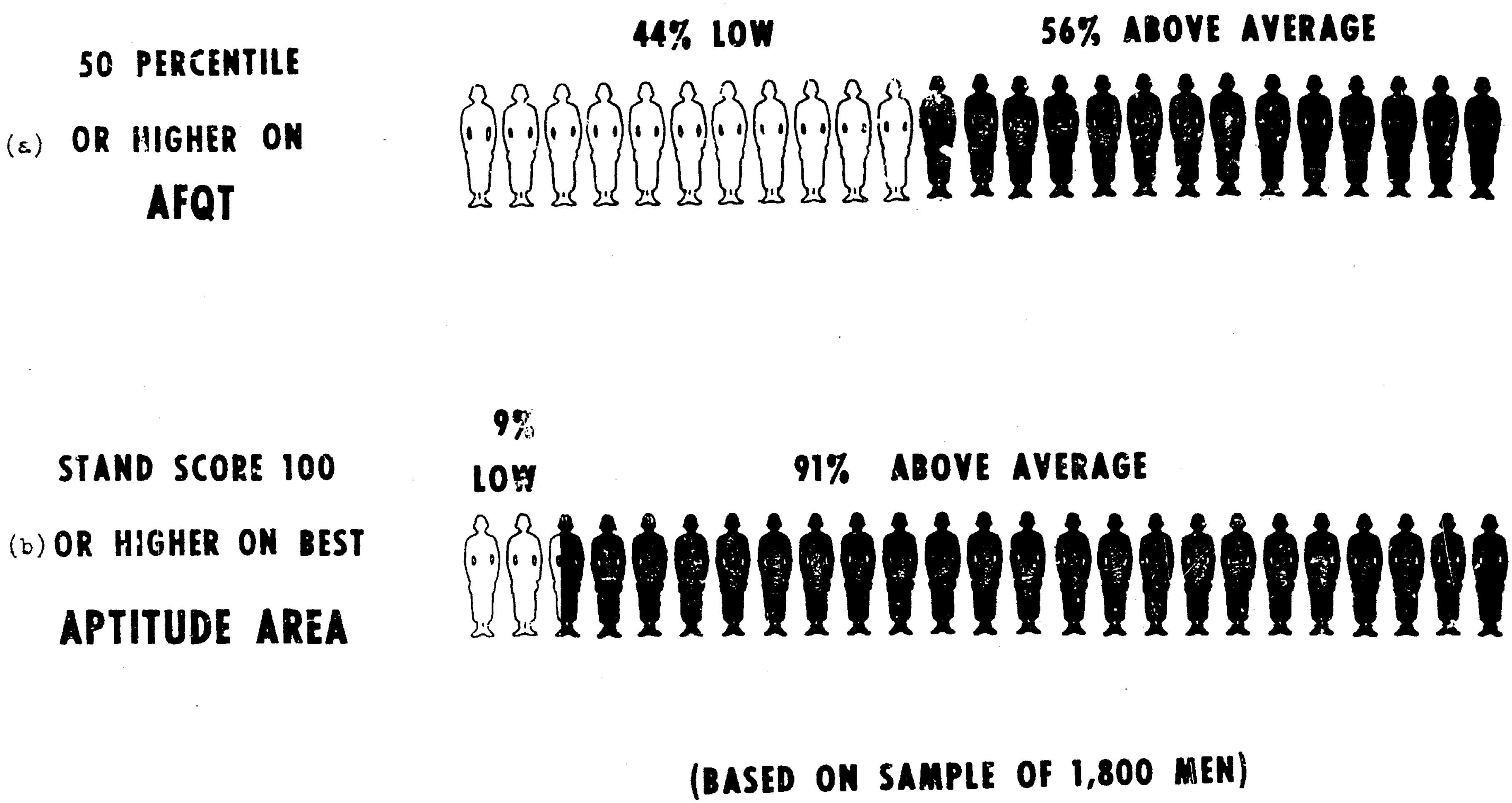


Figure 2. Proportion of sample of 1800 from input population of enlisted men scoring above 50th percentile (a) on AFQT and, (b) on their highest aptitude area.

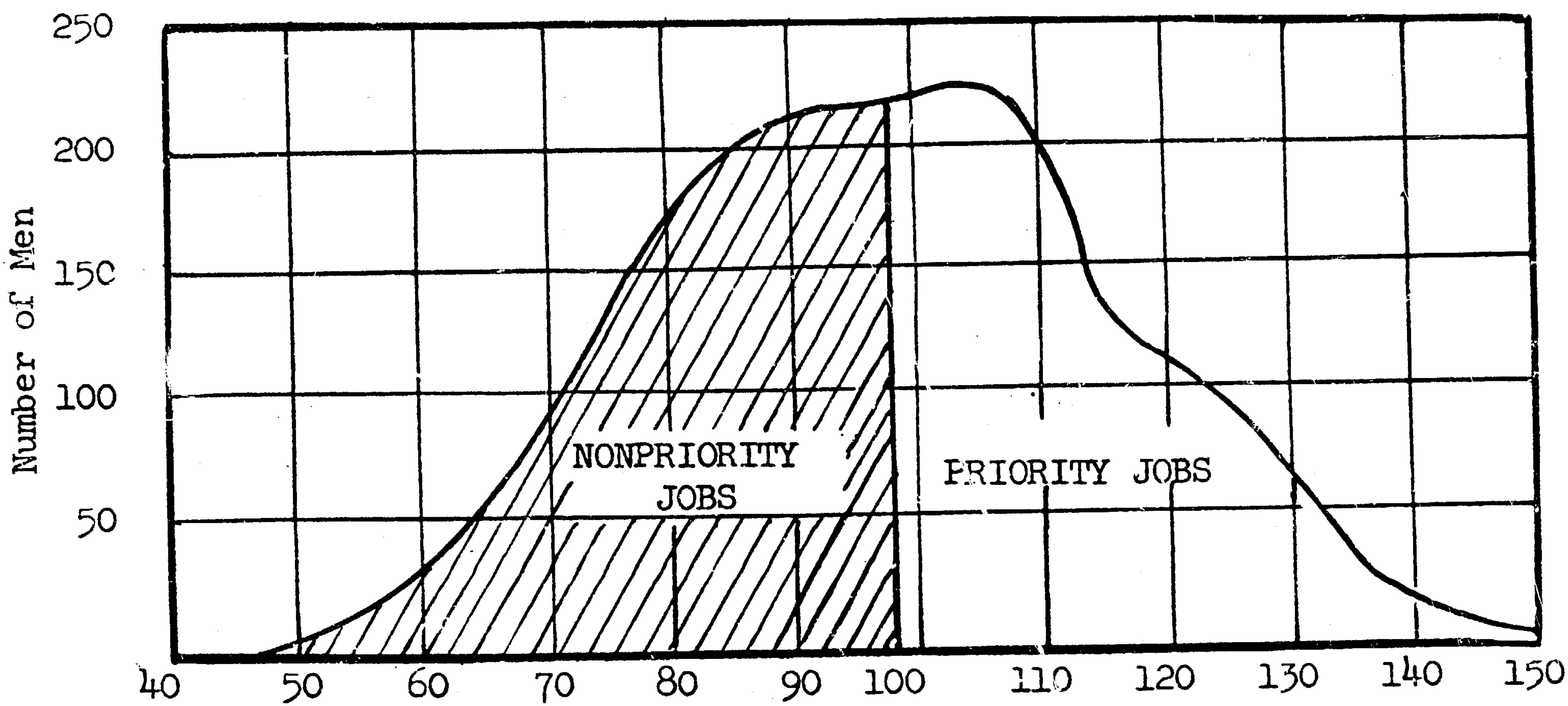


Figure 3. Distribution of Army Standard Scores on overall general ability for priority and nonpriority jobs, when assignment is based on a single measure.

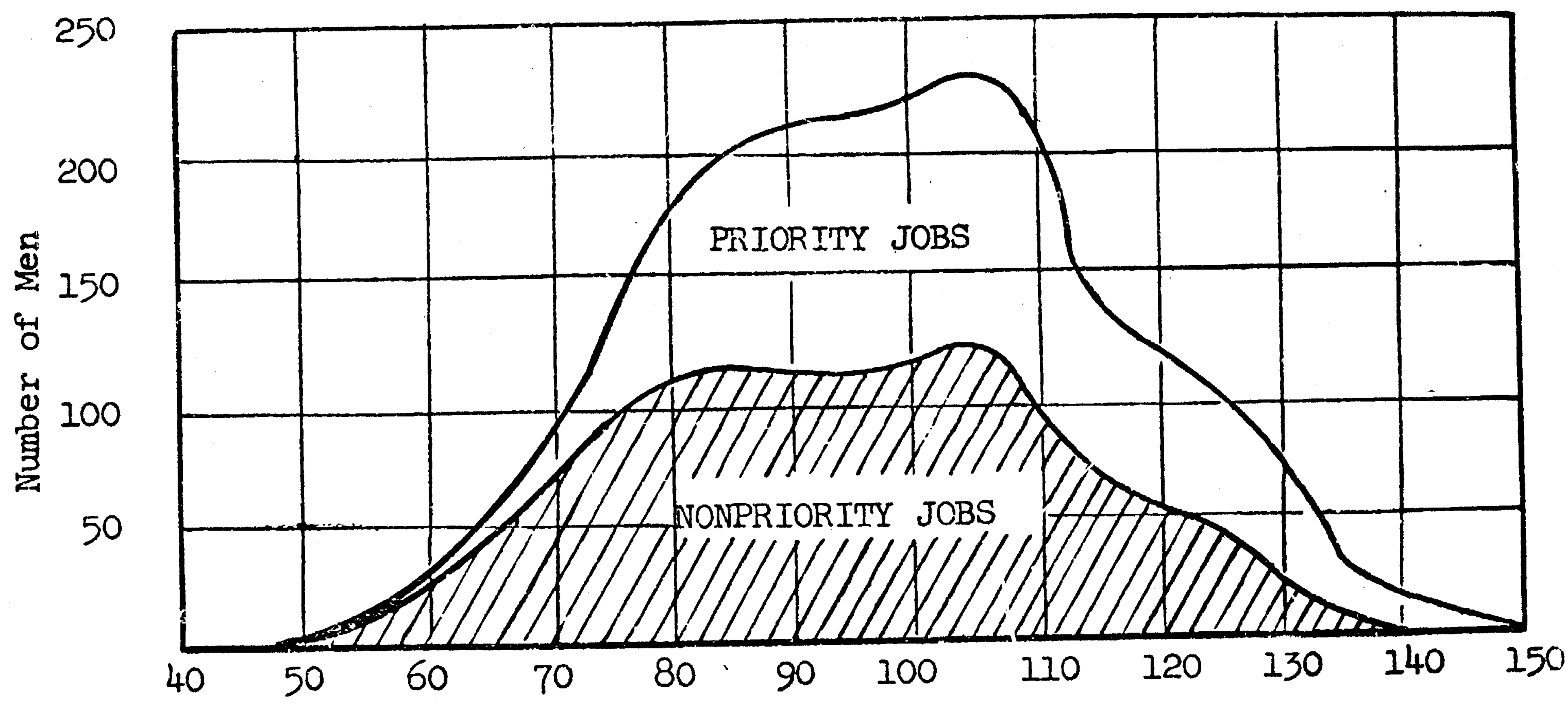


Figure 4. Distribution of Army Standard Standard Scores on overall general ability when assignment is based on battery of tests.

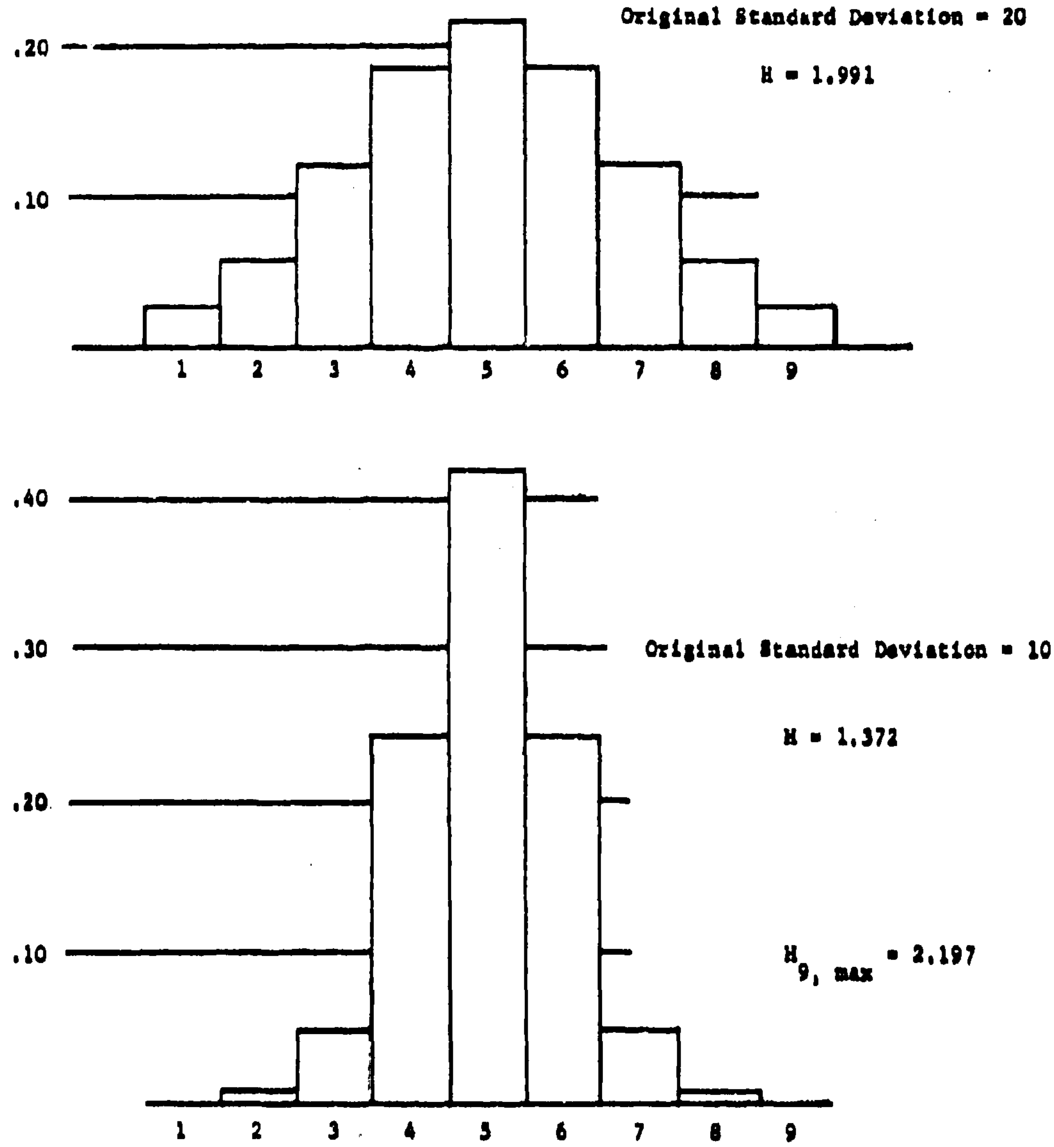


Figure 5. Discrete distributions resulting from continuous distributions with standard deviation of 10 and of 20.

	16 Variables		4 Variables	
Job	Natural	Perturbed	Natural	Perturbed
1	.0625	.0062	.2500	.1141
2	.0625	.0137	.2500	.2047
3	.0625	.0212	.2500	.2953
4	.0625	.0287	.2500	.3859
5	.0625	.0362		
6	.0625	.0437		
7	.0625	.0512		
8	.0625	.0587		
9	.0625	.0662		
10	.0625	.0737		
11	.0625	.0812		
12	.0625	.0887		
13	.0625	.0962		
14	.0625	.1037		
15	.0625	.1112		
16	.0625	.1187		

Figure 6. Job quotas, expressed as proportions, imposed as constraints on personnel assignment procedure.

Factorial Design for Experiment Using Simulated Performance Estimates

Factor 1: Standard deviation
- Level a: S = 10
- Level b: S = 20

Factor 2: Number of cases
- Level a: N = 160
- Level b: N = 320

Factor 3: Number of variables
- Level a: V = 4
- Level b: V = 16

Factor 4: Number of intervals
- Level a: I = 9
- Level b: I = 99

Factor 5: Quota
- Level a: Perturbed quotas
- Level b: Natural quotas

Figure 7. Experimental conditions used in the five-factor experiment.

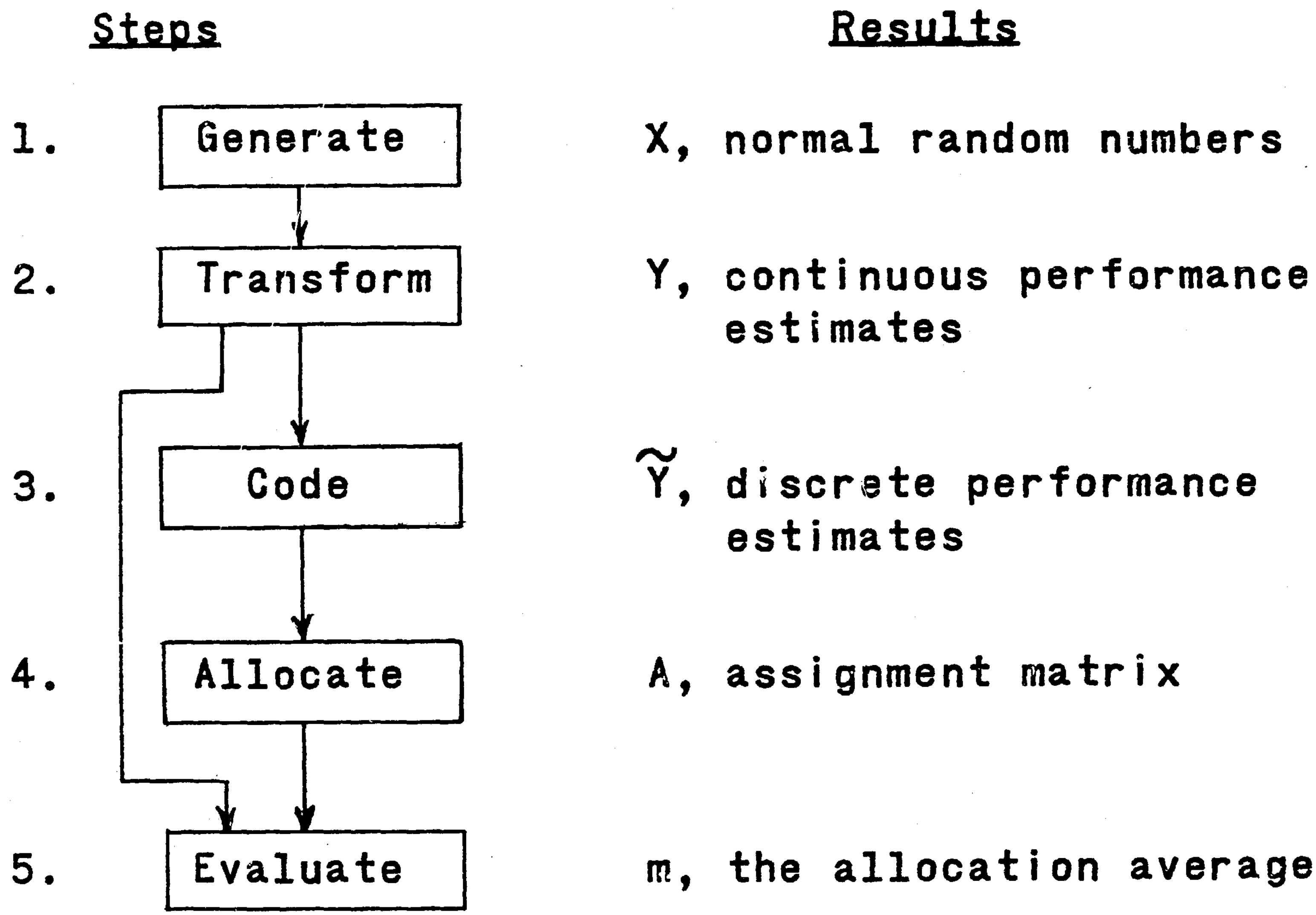


Figure 8. Flow diagram of experiment using simulated performance estimates.

Results of Experiment Using Simulated Performance Estimates

Factor	Level	Mean Performance: Standard Units	Mean Performance: Army S. Scores
Standard deviation	10	.65	113
Standard deviation	20	.70	114
Nr. of variables	4	.50	110
Nr. of variables	16	.86	117
Nr. of score intervals	9	.62	112
Nr. of score intervals	99	.74	115
Quota	Perturbed	.67	114
Quota	Natural	.69	114

Figure 9. Mean performance for selected factors.

Standard Deviation

		10	20
Number of intervals	9	.58 [111.5]	.67 [113.4]
	99	.73 [114.7]	.74 [114.8]

Entries are in terms of standard units; bracketed values are in terms of Army Standard Scores.

Figure 10. Mean performance for selected cells: First order interaction terms for standard deviation and number of intervals.

Number of intervals	Number of Variables: 4	Number of Variables: 16
9	.46 [109.2]	.78 [115.7]
99	.55 [110.9]	.93 [118.5]

Entries are in terms of standard units; bracketed values are in terms of Army Standard Scores.

Figure 11. Mean performance for selected cells: first order interaction terms for number of variables and number of intervals.

FACTORIAL DESIGN FOR EXPERIMENT USING SIMULATED PSYCHOLOGICAL TEST VARIABLES

Factor 1: Test intercorrelation

Level a: $r_{ij} = .4$

Level b: $r_{ij} = .6$

Factor 2: Test reliability

Level a: $r_{tt} = .7$

Level b: $r_{tt} = .9$

Factor 3: Number of items

Level a: $n = 20$

Level b: $n = 40$

Factor 4: Score conversion

Level a: Score

Level b: Stanine

Figure 12. Experimental conditions used in split-level experiment.

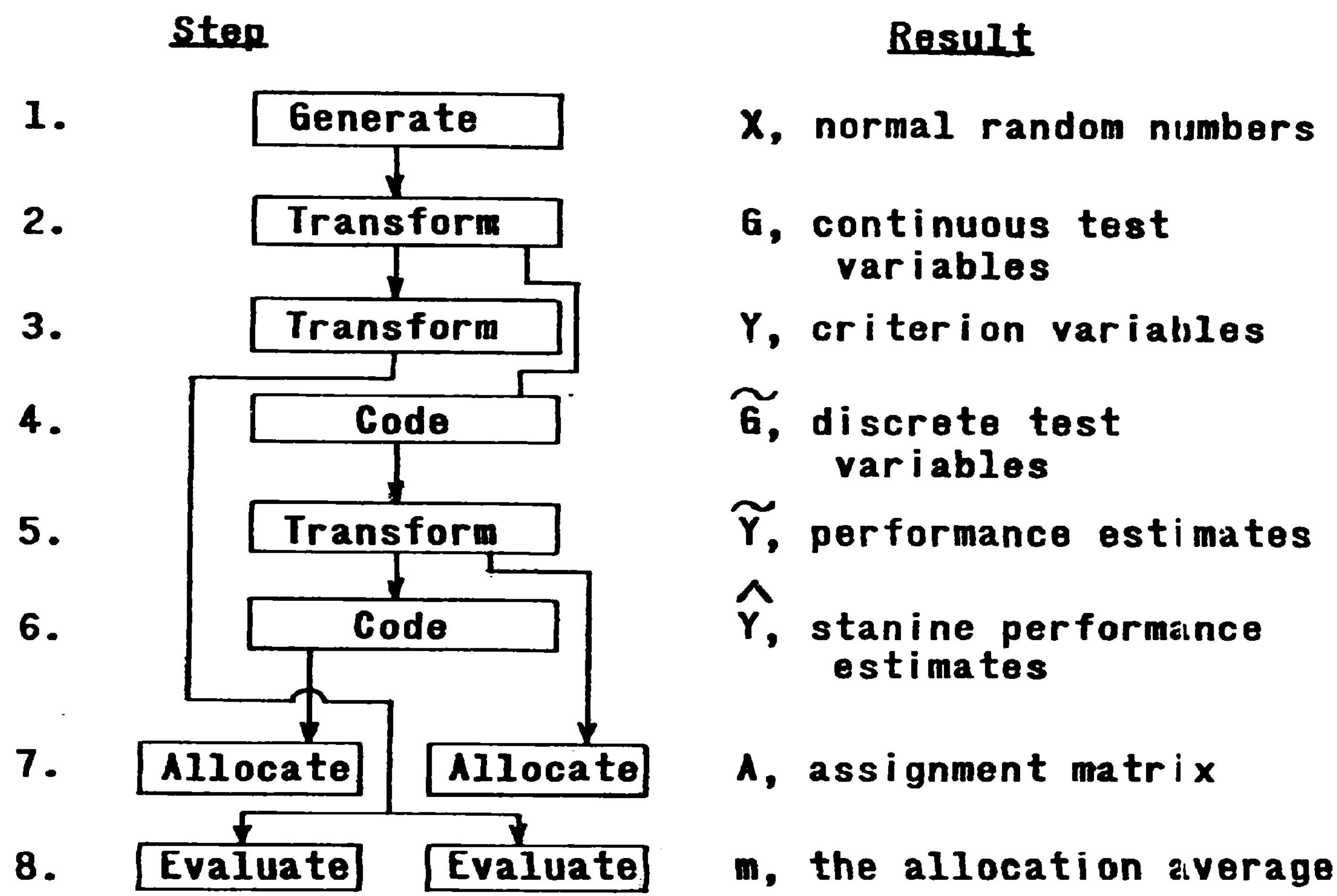


Figure 13. Flow diagram of experiment using simulated psychological test variables.

Source of Variation	Sum of Squares	d.f.	Mean Square	F	P
Intercorrelations	6.2960855	1	6.2960855	784.97	.001
Reliability	5.8434368	1	5.8434368	728.54	.001
Nr. of Items	.0133628	1	.0133628	1.67	.25
Intercorrelations x Reliability	.3419087	1	.3419087	42.63	.001
Intercorrelations x Nr. Items	.0087315	1	.0087315	1.09	---
Reliability x Nr. of Items	.0228176	1	.0228176	2.84	.10
Intercor. x Reliability x Nr. Items	.0002393	1	.0002393	---	---
Samples in Same Exp. Condition	.5774944	72	.0080208	---	---
Total between Samples	13.1040766	79			
Total within Samples	.0560724	80			
Total	13.1601490	159			

Figure 14. Results of analysis of variance applied to the average performance for independent variables: intercorrelation, reliability and number of items (between sample variance).

ANALYSIS OF VARIANCE OF ALLOCATION AVERAGE

Source of Variation	Sum of Squares	d.f.	Mean Squares	F	P
Score Conversion	.0441774	1	.0441774	777.78	.001
Score x Intercorrelation	.0000479	1	.0000479	---	---
Score x Reliability	.0061563	1	.0061563	108.39	.001
Score x Nr. of Items	.0001583	1	.0001583	2.79	.10
Score x Intercor. x Reliability	.0003036	1	.0003036	5.35	.025
Score x Intercor. x Nr. of Items	.0000131	1	.0000131	---	---
Score x Reliability x Nr. of Items	.0011273	1	.0011273	19.85	.001
Score x Intercor. x Reliab. x Nr. Items	.0000003	1	.0000003	---	---
Sample x Score	.0040882	72	.0000568	---	---
Total within Samples	.0560724	80			
Total between Samples	13.1040766	79			
Total	13.1601490	159			

Figure 15. Analysis of variance of average performances for the correlated variable: type of score conversion (within sample variance).

Results of Experiment Using Simulated Test Scores

Factor	Level	Mean Performance: Standard Units	Mean Performance: Army S. Scores
Test intercorrelation	.4	.83	117
Test intercorrelation	.6	.43	109
Test reliability	.7	.45	109
Test reliability	.9	.82	116
Nr. of items	20	.62	112
Nr. of items	40	.64	113
Score conversion	Score	.65	113
Score conversion	Stanine	.61	112

Figure 16. Mean performance for the four factors.

INTERACTION TERMS

Test intercorrelation x test reliability

		Intercorrelation .4	Intercorrelation .6
Reliability	.7	.68 [113.7]	.20 [103.9]
Reliability	.9	.97 [119.5]	.67 [113.4]

Entries are in terms of standard units; bracketed values are in terms of Army Standard Scores.

Figure 17. Mean performance for selected cells: first order interaction terms for intercorrelation and reliability factors.

ANALYSIS OF VARIANCE OF ALLOCATION AVERAGE

(30 Replications)

Source of Variation	Sum of Squares	d.f.	Mean Square	F	P
Number of Items	.0727871	1	.0727871	6.77	.025
Samples in Same Exp. Cond.	.6237790	58	.0107548	----	----
Between Samples	.6310577	59			
Score Conversion	.0179231	1	.0179231	25.21	.001
Score x Number of Items	.0009848	1	.0009848	1.39	.25
Score x Sample	.0412308	58	.0007109	----	----
Within Samples	.0601387	60			
Total	.6911964	119			

Figure 18. Results of analysis of variance contrasting number of items at a fixed level of reliability and intercorrelation (.7 and .4, respectively).

Means for Experiment Using Additional Replications

Nr. of Items

	20	40	Total
Score	.676 [113.5]	.731 [114.6]	.703 [114.1]
Stanine	.657 [113.1]	.701 [114.0]	.679 [113.6]
Total	.667 [113.3]	.716 [114.3]	.691 [113.8]

Entries are in terms of standard units; bracketed values are in terms of the equivalent Army Standard Scores.

Figure 19. Mean performance for replicated cells. (reliability = .7, intercorrelation = .4)

QUANTITATIVE ASSAY FOR CRUDE ANTHRAX TOXINS*

Bertram W. Haines, Frederick Klein, and Ralph E. Lincoln
U. S. Army Biological Laboratories
Fort Detrick, Frederick, Maryland

ABSTRACT. The whole crude toxins of *Bacillus anthracis*, although apparently responsible for the death of animals with anthrax, had never been quantitated. A total of 14 lots of the toxic culture filtrate of *B. anthracis* were pooled into one large lot of crude anthrax toxins. An extensive assay of this reference material was conducted in four laboratories by use of the time-to-death of the intravenously challenged Fischer 344 rat as the response variable. Doses of the material were varied factorially by concentration, dilution, and volume. The data from this study were used to define a potency unit of the crude anthrax toxins. Procedures were developed and illustrated for the assay of unknown lots of the toxins by comparing the rate time-to-death response to the unknown with either (i) the responses reported in this study, or (ii) directly with the rat responses to a new sample of the reference toxins. The possibilities and limitations of this standardization and of the statistical procedure through which it was developed are discussed.

INTRODUCTION. The excellent work of Smith, Keppie, and Stanley (1955a), demonstrating the toxins of *Bacillus anthracis* organisms in the blood from guinea pigs in the terminal stages of anthrax, rekindled interest in the disease, particularly its toxins. (The toxic methabolic by-products of the growth of *B. anthracis* are composed of components with different biological or chemical properties. Naturally produced combinations of these components in unknown proportions will be referred to in this paper as "toxins.") To date, valid comparisons of results among the several experimenters (Smith et al., 1955a, b, 1956; Smith and Gallop, 1956; Thorne, Molnar, and Strange, 1960; Stanley and Smith, 1961; Beall, Taylor, and Thorne, 1962; Klein et al., 1962; Keppie, Smith, and Harris-Smith, 1955; Eckert and Bonventre, 1963; Harris-Smith, Smith, and Keppie, 1958; Sargeant, Stanley, and Smith, 1960; Stanley, Sargeant, and Smith, 1960) who have reported work with the toxic materials produced by *B. anthracis* have been difficult, because either whole crude toxins or the several components have been assayed by different methods, in different assay animals, and with no reference standard of the toxins.

---

*This paper appeared previously in the Journal of Bacteriology, Volume 89, No. 1, pages 74-83. Permission of the editors of this journal to publish this paper in these Proceedings is appreciated.

This paper presents the results of studies to quantitate, in terms of defined potency units, the lethality of anthrax toxins in Fischer 344 rats. The authors developed a reference lot of stabilized freeze-dried crude anthrax toxins. This reference material was used in the study described here, and is available for other studies against which samples of anthrax toxins of unknown concentration can be assayed.

MATERIALS AND METHODS. Animals. Fischer 344 albino rats weighing 200 to 300 g were obtained from the Fort Detrick colonies of Frank Beall and Frederick Klein. Both colonies are maintained through brother-sister matings descended from the colony described by Taylor, Kennedy, and Blundell (1961). This weight range was chosen, because preliminary data indicated that the response time of rats that weigh more than 300 g was significantly greater than that of rats weighing more than 200, but less than 300, g. Further study on rats, carefully selected for weight, revealed no significant difference within the weight range of 200 to 300 g (Table 1). The analysis of variance is presented in Table 2.

TABLE 1
Response time in minutes of 27 rats injected with 1 ml of crude anthrax toxins by weight of rat.

Weight (g) of rat		
200	250	300
99	102	100
97	81	94
96	80	88
94	79	105
93	78	90
92	114	101
89	76	78
88	102	82
87	71	86
835*	783	824
92.6**	84.9	90.7

* Totals
** Harmonic means.

TABLE 2
Analysis of variance of reciprocal response times recorded in Table 1

Source	df*	Sum of squares	Mean square	F
Between weights	2	.0485	.0242	1.50**
Within weights	24	.3859	.0161	
Total	26	.4344		

* Degrees of freedom.
** Not significant.

Rat lethal test. Toxins of B. anthracis were injected into the dorsal vein of the penis of the Fischer rat. In describing this test, Beall et al. (1962) noted a definite relationship between the dose of the toxins injected and time-to-death.

Antiserum. Equine hyperimmune serum (DH-1-6C) prepared by repeated injections of spores of the Sterne strain of B. anthracis, was used (Thorne et al., 1960).

Preparation of anthrax toxins. The medium used was described by Thorne et al. (1960), and was made with triple-distilled water. Subsequent to his original description, Thorne (personal communication) has suggested some changes. The medium used in this study was as follows.

Nine stock solutions (A, B, C, D, E, F, G, H, and I) were prepared. All stock solutions may be stored at 4 C for indefinite periods of time. Solution A contained $CaCl_2 \cdot 2H_2O$, 0.368 g/500 ml of water; B contained $MgSO_4 \cdot 7H_2O$, 0.493 g/500 ml of water; C contained $MnSO_4 \cdot H_2O$, 0.043 g/500 ml of water; D contained adenine sulfate, 0.105 g, and uracil, 0.070 g (both solids were dissolved in 100 ml of water, and the total volume was made up to 500 ml).

Solution E contained thiamine HCl, 0.025 g/500 ml of water; F contained tryptophan, 2.600 g; cystine, 0.600 g; and glycine, 0.750 g. The solids in solution F were dissolved as follows. Tryptophan was dissolved in 6 ml of 6 N HCl. Cystine was dissolved in 100 ml of water. Glycine was dissolved in 150 ml of water. These three solutions were combined, and water was added to bring the total volume up to 500 ml.

Solution G contained KH_2PO_4, 34.0 g/500 ml of water; H contained K_2HPO_4, 43.6 g/500 ml of water; I contained charcoal (Norit A), 3.75 g/500 ml of water.

A 10-ml amount of each stock solution, except that containing charcoal, was added to a suitable container; and 3.6 g of Casamino Acids (Difco) were added. The volume was brought up to 1 liter with triple-distilled water, and the pH of the medium was adjusted to 6.9 with 1 N H_2SO_4 or 1 N NaOH as needed. A 460-ml amount of this preparation was dispensed into a 3-liter Fernbach flask; 2 ml of charcoal suspension were added, and the preparation was autoclaved for 20 min at 15 psi.

Inoculation procedure. A 5-ml amount of 20% glucose (sterilized by filtration) was added to the Fernbach flask containing 460 ml of sterilized basal medium. Each flask of final medium was inoculated with 2×10^6 Sterne strain spores. The inoculated flasks were incubated statically for 23 to 27 hr at 37 C; 4 hr after inoculation 55 ml of 9% $NaHCO_3$ were added to each flask.

This final culture was centrifuged at 3,000 X g for 30 min. The supernatant fluid was decanted, and 10% horse serum was added. The solution was then sterilized by filtration through an ultrafine glass filter.

A preliminary test, to determine the potency of each of 14 toxic filtrates, was done by injecting 1-ml samples of each filtrate intravenously into two rats. The response (death) times of the rats were considered as indications of the toxicity of each batch. The total volume per batch and the response times of the test rats are given in Table 3.

The 14 toxic filtrates were combined, and a second preliminary test was conducted on the pooled material. The two rats used in this test died in 104 and 117 min, with a mean response time of 110.5 min. Both response times are within one standard deviation of the mean of all batches.

The pooled toxins were dispensed into 600 drying ampoules (40 ml), each containing 10 ml of toxins. Ampoules were shell-frozen in Dry Ice and alcohol (-79 C). Frozen ampoules were placed on an Aminco Dryer (American Instrument Co., Silver Spring, Md.), and dried under vacuum

TABLE 3
Volume per batch and response time of rats challenged with toxins by batch

Batch	Total volume ml	Response time (min)		
		Rat A	Rat B	Mean
1	450	97	92	94.5
2	450	107	91	99.0
3	450	97	96	96.5
4	460	95	—*	95.0
5	420	122	124	123.0
6	450	114	125	119.5
7	510	116	90	103.0
8	410	121	120	120.5
9	370	88	82	85.0
10	510	90	94	92.0
11	465	106	94	100.0
12	425	106	92	99.0
13	425	117	121	119.0
14	300	100	117	108.5
Total	6,095			103.9**

* Missed the vein.
** SD = 12.14.

of 10 to 30 μ of mercury for 18 to 25 hr. Ampoules were sealed under vacuum, packed in cardboard containers, and stored at -20 C. A third preliminary test was conducted at this point. One randomly selected ampoule was reconstituted with 10 ml of triple-distilled water. A 1-ml amount of this toxic material was assayed in each of five rats. Their mean response time was 117.2 min. To further test the toxicity, 0.2 ml of undiluted and of serial twofold dilutions of the reconstituted material was injected intradermally into the shaven sides of a guinea pig, and observed for edematous reaction. The material reacted at a dilution of 1:32, and can be expressed according to Thorne et al. (1960) as containing 32 toxic units. Additional vials were reconstituted to 4X concentration, and tested on immunodiffusion plates against the standard spore antiserum (Thorne et al., 1960). Three individual lines of precipitate appeared in parallel arrangement when tested with a linear pattern. The strongest

precipitate line was identified as the protective antigen (factor II) component when compared with a standard (Beall et al., 1962). An undiluted sample of the resuspended material had a protective antigen titer of 1:64 against the standard spore antiserum.

Reference toxins. These preliminary tests constituted quality control measures on the remaining 597 vials of dried toxic filtrate. As a result of these tests, it was known that these vials contained the known components of anthrax toxins.

Procedures. The toxins were assayed independently by each of four investigators. The procedures followed by each of the four were as similar as possible.

The characterization of the dose-response relationship of the toxins in Fischer rats was based on an assay in which the two dose factors of amount and concentration of toxins were each tested at several levels as follows: (i) five levels of the amount of toxins designated as 4 ml, 2 ml, 1.5 ml, 1 ml, and 0.5 ml; (ii) seven levels of the concentration of the toxins designated as 4X, 2X, 1X, 0.5X, 0.25X, 0.125X, and 0.0625X, where 1X is defined as the concentration resulting when 1 ampoule is reconstituted to 10 ml with a diluent of triple-distilled water. Dilutions beyond 1X were made with distilled water plus 10% normal horse serum.

The 7 X 5 factorial combinations of the several levels of these two factors, plus 19 control groups, were each tested in two Fischer rats by each of four investigators (Table 4). Three sets of control animals are not shown in Table 4. The first set included five pairs of rats. Each pair was inoculated with one of the five amounts of diluent along (i.e., triple-distilled water plus 10% normal horse serum) to provide assurance that their companion animals responded to toxins as opposed to the inoculation of the diluents. The second set included seven pairs of animals. Each pair in this set was inoculated with 1.5 ml of one of the seven concentrations of toxins mixed with 0.5 ml (1/3 by volume) of specific antiserum (Thorne et al., 1960). The seven pairs of animals in the third set of controls were inoculated with 1.5 ml of one of the seven concentrations of toxins mixed with 0.5 ml of normal horse serum. These animals provided assurance that the control no. 2 animals that lived were saved by the antiserum specific against anthrax toxins.

TABLE 4
Response times in minutes of 280 Fischer rats by dose, concentration, technician, and rat

Concn	Tech-nician	4*		2*		1.5*		1*		0.5*	
		A**	B	A	B	A	B	A	B	A	B
4X	1	58	55	53	54	57	57	61	60	76	71
	2	53	61	54	52	64	63	64	63	85	70
	3	57	62	56	52	58	56	64	62	78	72
	4	60	52	448	53	59	123	63	59	81	82
2X	1	57	57	61	63	59	61	72	70	100	89
	2	57	55	65	62	74	65	84	77	119	94
	3	50	56	56	58	66	77	72	78	109	117
	4	67	56	55	65	67	S***	127	S	107	83
1X	1	53	55	70	69	119	70	90	91	127	159
	2	73	64	78	72	82	81	61	100	181	199
	3	65	62	77	80	89	83	107	97	293	483
	4	S	63	S	S	S	100	132	S	161	202
0.5X	1	70	77	153	143	129	134	145	148	S	S
	2	74	83	114	103	138	131	425	281	S	S
	3	75	69	113	118	137	151	1588	244	S	S
	4	74	94	S	139	149	S	S	400	S	S
0.25X	1	111	112	173	176	S	481	S	S	S	S
	2	136	176	295	274	S	S	S	S	S	S
	3	103	124	S	300	S	S	S	S	S	S
	4	S	118	S	S	S	S	S	S	S	S
0.125X	1	185	195	S	S	S	S	S	S	S	S
	2	253	588	S	S	S	S	S	S	S	S
	3	473	234	S	S	S	S	S	S	S	S
	4	S	S	S	S	S	S	S	S	S	S
.0625X	1	S	S	S	S	S	S	S	S	S	S
	2	S	S	S	S	S	S	S	S	S	S
	3	S	S	S	S	S	S	S	S	S	S
	4	S	S	S	S	S	S	S	S	S	S

* Dose expressed in milliliters.
** Rat A or B.
***S indicates survival.

Each investigator required 32 ampoules of dried toxins. Each of the 32 ampoules was opened, and reconstituted with 2.5 ml of diluent precooled to 4 C. The contents of all 32 ampoules were then pooled, providing a total of 80 ml of reconstituted toxins at a concentration of 4X (4 times the original). All concentrations of toxins were maintained continuously at 4 C. To make the next dilution, 40 ml of the pool (4X) were combined with 40 ml of diluent (triple-distilled water). This provided 80 ml of toxins at a concentration of 2X. Further serial twofold dilutions were made to 0.0625X (1/16 X original concentration) and inoculated as planned.

Each investigator required 108 rats. These rats were caged in 54 consecutively numbered cages, each containing two animals. Each of the 54 treatment combinations was given to the two animals in one cage at the same time. The order of the treatments was randomized for each investigator. Response times-to-death, in minutes, were recorded for each rat and constituted the basic data.

RESULTS. The response times for animals are presented in Table 4. Although none of the controls appears in this table, none of either the first or second groups of control animals died. Some animals in the third control group challenged with 1.5 ml of toxins plus normal horse serum responded nearly the same as test animals challenged with 1.5 ml of toxins. The mean response times, in minutes, of these control animals by concentration of toxins are recorded in Table 5. The pattern of responses by the controls provided the needed assurance that the response of the test animals was specifically to the toxins of B. anthracis.

TABLE 5
Mean response time by dose and concentrations of toxins

Concn of toxin	Dose (ml) 4	2	1.5	1	0.5	Mean	Control*
4X	57.5	53.5	59.0	62.3	75.0	60.7	60.0
2X	55.2	60.7	66.4	75.2	105.1	69.0	70.0
1X	61.3	74.1	85.1	88.0	198.7	86.3	134.0
0.5X	74.4	121.6	136.3	247.0	S**	151.3	154.0
Mean	61.3	70.3	78.3	89.4	143.5	91.3	

*Control was 1.5 ml of toxins plus normal horse serum.
**All animals survived.

In spite of carefully controlled procedures and techniques, the results from one laboratory (technician 4) were so erratic that they were disregarded in any further analysis. Inspection of these data showed that technician 4 was the only one having reversal of results; i.e., a greater amount of toxins not killing and lesser amounts killing, or only one of the two test animals responding (except at doses eliciting a response above 300 min). These extremely variable results indicated that adequate controls on technique and environment were not maintained in this laboratory.

The reciprocals of the response times were used for analysis, because reciprocal response times are nearly normally distributed with equal variances, whereas the untransformed response times are positively skewed with unequal variances (Finney, 1952). The analysis of variance on the reciprocal response times of 120 rats from the four highest concentrations and the five doses is shown in Table 6. From this analysis it was seen that both dose level and concentration had statistically significant effects on the response time of Fischer rats injected intravenously with anthrax toxins.

TABLE 6
Analysis of variance of reciprocal response times

Line no.	Effect	df	Sum of squares	Mean square	F*
1	Dose (D)	4	11.9272	2.9818	229.37·*
2	Concentration (C)	3	16.5629	5.5210	424.69**
3	Technician (T)	2	0.1543	0.0772	5.94***
4	D X C	12	1.7984	0.1499	11.53**
5	D X T	8	0.1485	0.0186	1.43
6	D X T	6	0.1180	0.0197	1.52
7	D X C X T	24	0.6452	0.0269	2.07
8	Error	60	0.7814	0.0130	
9	Total	119	32.1360		

* Error line 8 was used to test all effects.
** Approximate probabilities < 0.001.
***Approximate probabilities < 0.05.

The analysis further showed an interaction between dose and concentration to be statistically significant. The mean response times by does and concentration of toxins are given in Table 5. From the tabled means, it can be seen that the magnitude of this interaction is slight and had no practical significance in the further analysis and interpretation of these data.

The analysis also showed a statistically significant difference among technicians. Inspection of the data showed that mean response times for all rats responding for technicians 1, 2, and 3 were, respectively, 78, 83, and 83 min. This is a practically unimportant difference which we believe may in part be due to environmental factors, because genetic differences would be almost nil after 100 generations of inbreeding. The rats used by technician 1 came from the Beall colony, which was maintained in a different environment than the Klein colony animals used by the other two technicians. This raised the question as to the effect on this assay of Fischer rats procured from non-Detrick sources. To examine this effect, commercially available Fischer rats obtained from two breeders were tested and found to be suitable for this assay. In this study, 20 Fischer 344 rats from each of two suppliers (Microbiological Associates, Inc., Bethesda, Md.; and Charles River Breeding Laboratories, Inc., Brookline, Mass.) were challenged in each of two laboratories. The response times of all 80 rats are reported in Table 7. No statistically significant difference in times of response for animals from the two suppliers was observed. A difference between the two operators and the interaction of operator X supplier was statistically significant at the 5% level. The mean response time of three of the four groups differed by less than 1 min, and the fourth group differed by approximately 5 min. This difference of about 5 min between these two groups could be caused by a difference of about seven units of toxins, which is well within the 95% confidence limits of an estimated potency. Thus, this difference, although statistically significant, was considered of no consequence concerning this assay.

A test to determine the storage characteristics of the reference toxins was conducted on a vial of the toxins which had been stored for 36 months. The test vial was reconstituted with 10 ml of triple-distilled water. Six rats were then challenged with these reconstituted toxins, according to the protocol described in this paper.

TABLE 7
Response times in minutes by supplier, operators, and rats

Rats	Charles River Breeding Labs., Inc.		Microbiological Associates, Inc.	
	1*	2	1	2
1	83	87	91	85
2	88	84	84	89
3	86	86	91	89
4	83	82	88	85
5	91	84	89	92
6	87	89	88	84
7	94	88	90	101
8	88	83	92	87
9	87	83	96	102
10	91	86	77	87
11	105	83	89	93
12	94	85	94	79
13	92	79	90	107
14	90	81	91	88
15	98	81	91	83
16	91	85	77	90
17	82	83	97	89
18	90	87	89	88
19	83	85	82	75
20	88	83	90	86
Harmonic mean response time	89.28	84.10	88.50	88.42

*Operator number.

The estimate of potency from that test was 32.4 potency units per ml at the 1X concentration. This was essentially identical to the 32 units per ml set up in the definition. Therefore, it was concluded that the reference toxins had not changed with respect to potency during 36 months of storage.

Development of procedures for direct assay method. A potency assay should be based on dose expressed in terms of well-defined units. No such units have as yet been defined for anthrax toxins. Varying the amount of toxins by varying either dose or concentration would have a significant effect on the response time of rats; however, rats injected with 1 ml of toxins concentrated to 2X responded in about the same time (75 min) as rats injected with 2 ml of toxins concentrated at 1X (74 min). This relationship holds true for most other dose-by-concentration combinations for which the product of these two factors is a constant. If doses are converted into 0.5-ml units, and concentrations into 0.0625 units, then the doses and concentrations in Table 4 can be expressed as shown in Table 8.

TABLE 8
Derivation of potency units of anthrax toxins

Concn of toxins in 0.0625-fold units	Dose of toxins in 0.5-ml units				
	8	4	3	2	1
64	512	256	192	128	64
32	256	128	96	64	32
16	128	64	48	32	16
8	64	32	24	16	8
4	32	16	12	8	4
2	16	8	6	4	2
1	8	4	3	2	1

The products of the marginal numbers in Table 8 for any two equivalent dose-by-concentration combinations are the same; thus, the product of two dose units and 32 concentration units gives 64 total potency units of toxins. Similarly, four dose units of 16 concentration units also contain 64 total potency units of toxins. We define the potency unit of anthrax toxins to be expressed as these products of dose by concentration of this particular lot of toxins.

If we were to carry the definition of a potency unit no further, then 1 ml of 1X concentration of any anthrax toxins, regardless of its actual effect in animals, would have 32 potency units. To standardize a potency unit, it is necessary to describe the association between the dose, in units, and the potency, in terms of a biological response to this particular lot of anthrax toxins. The potency of any other lot of toxins may then be measured by comparing the response to a known amount of the test toxins with the response to the same amount of the reference toxins.

These response characteristics were described as the dose-response relationship when measured doses of these toxins were injected intravenously into Fischer 344 rats. The challenged rats responded by dying at a time that is shown here to be highly dependent on the dose measured in potency units of these toxins.

The regression of mean reciprocal response times on the $\log_2$ of the potency units of anthrax toxins is shown in Figure 1. The least squares line has the equation:

$$Y = b_0 + b_1X + b_2X^2 \tag{1}$$

where Y is the mean reciprocal response time, X is the potency of anthrax toxins in $\log_2$ units, and the b values are regression coefficients computed from the data of this test. The values of the coefficients, their variances and covariances, are: $b_0 = -2.591$; $b_1 = 0.959$; $b_2 = -0.051$; $V(b_0) = 0.077121$; $V(b_1) = 0.009514$; $V(b_2) = 0.000068$; $V(b_0b_1) = -0.026902$; $V(b_0b_2) = 0.002238$; $V(b_1b_2) = -0.000800$. This regression line represents a basis upon which comparisons of potency of anthrax toxins can be made. Thus, test toxins can be assayed either indirectly against this curve, or directly with parallel assays of the reference toxins.

<u>Development of procedures for indirect assay method</u>. To use the responses of 120 rats to the reference toxins [for which the slope of response from the regression data (Figure 1) has been calculated], we recommend use of the indirect method for standardizing unknown potencies of anthrax toxins. The regression was nearly linear for

doses from 16 to 128 units, corresponding to response times from 240 to 65 min. Thus, although the concentration of test or unknown toxins is arbitrary, it should be of such concentration that 1 ml, injected intravenously, will kill a Fischer rat in not less than 65 min, nor more than 240 min.

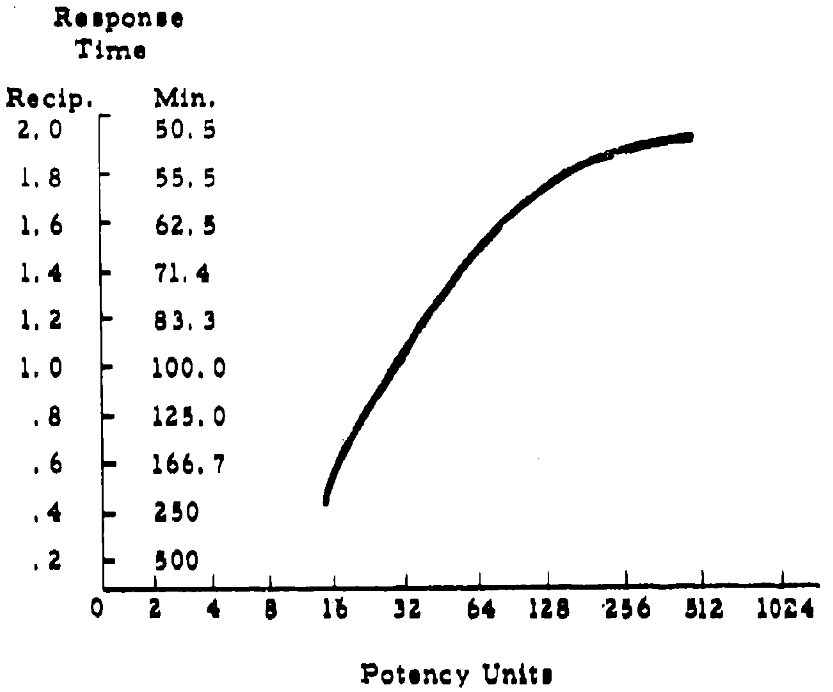


Figure 1. Regression of reciprocal response time of Fischer rats on log dose of anthrax toxins expressed in potency units.

To test the potency of test or unknown toxins, enough animals should be used so that the amount of variation in the final result, that can be attributed to the test rats, is at least no greater than the amount of variation contributed by the standard rats. Thus, at least six Fischer rats of 200 to 300 g from a suitable colony should be intravenously inoculated, three with 2 ml of the test toxins, and three with 1 ml.

The test is based on the mean reciprocal response times of the rats. (The rat response is very uniform; thus, any observed nonresponse must be considered the result of technique at some stage of the assay procedure.) This is simply the sum of reciprocal times-to-death of the rats in minutes (100/t) with the average time calculated. The reciprocal response times of the rats can be put in the following form:

Reference Toxins

$Y = 100/t$

	1 ml		2 ml	
Rat	1.	______	4.	______
	2.	______	5.	______
	3.	______	6.	______
	ΣY	______	ΣY	______
	$\bar{Y} = R_1$		R_2	______

$R_1 + R_2 =$ ______

Test Toxins

$Y = 100/t$

	1 ml		2 ml	
Rat	1.	______	4.	______
	2.	______	5.	______
	3.	______	6.	______
	ΣY	______	ΣY	______
	$\bar{Y} = T_1$	______	T_2	______

$T_1 + T_2 =$ ______

where R_1, R_2, T_1, and T_2 are mean reciprocal response times. This form for calculation can be used for either the direct or indirect assay method.

The estimate of the difference in potency (D) between the test toxins and the reference can be found as:

$$D = \frac{(T_1 + T_2) - (R_1 + R_2)}{2L} \qquad (2)$$

where the letters T and R represent the mean reciprocal response times from the table above, and L is the average slope of the reference dose-response curve at the two dose levels used in the test. This average slope may be calculated as:

$$L = b_1 + b_2 (X_1 + X_2) \tag{3}$$

where X_1 and X_2 are the dose levels of the reference toxins (in $\log_2$ potency units) that were used in the test, and b_1 and b_2 are the estimates of the regression coefficients from equation 1. When the test is run using 1- and 2-ml doses of toxins, then $X_1 = 5$ and $X_2 = 6$. Under these conditions $R_1 = 0.92$, $R_2 = 1.34$ from equation 1, and $L = 0.3985$ from equation 3, so that equation 2 becomes:

$$D = \frac{(T_1 + T_2) - 2.26}{0.7970} \tag{4}$$

where the letter D represents the amount of difference between the test and reference toxins in terms of $\log_2$ potency units. If D is positive, then the test toxins are more potent than the reference, whereas, if D is negative, the test toxins are less potent than the reference. The reference toxins have a potency of 5 $\log_2$ units per ml at a concentration of 1X; thus, the potency (P) of the test toxins in $\log_2$ units at the concentration tested will be found as:

$$P = 5 + D \tag{5}$$

To find the number of potency units per ml of the test toxins, its potency needs to be converted from $\log_2$ units to $\log_{10}$ units. The conversion formula is:

$$\log_{10} P = \log_2 P \log_{10} 2$$

The value of P in units is found by looking up the antilog of this product. This value will be the number of potency units per milliter of the test toxins at the concentration tested.

Estimation of variance. There is variation inherent in this assay system in addition to the variation between samples of toxins. Thus, the single estimates of the potency of any particular sample of an unknown toxin should be bounded by confident limits. To determine these limits it is necessary to calculate the variance (V) of the estimate D of the $\log_2$ of the difference in potency between the test and the reference. The variance of the estimate D will depend on the variances of the observed response times and of the regression.

If we express D as N/G where

$$N = (T_1 + T_2) - (R_1 + R_2) \tag{6}$$

and

$$G = 2L$$

then the variance of D can be expressed as:

$$V(D) = \frac{1}{4L^2} \left\{ V(N) + D^2 V(G) \right\} \tag{7}$$

which will apply, because N and G are estimated from independent observations (Finney, 1952). The four mean reciprocal response times are stochastically independent; thus, the estimate of V(N) can be expressed as:

$$V(N) = V(R_1) + V(R_2) + V(T_1) + V(T_2) \tag{8}$$

where $V(T_1)$ and $V(T_2)$ are obtained directly from the data of the test, and $V(R_1)$ and $V(R_2)$ are calculated from the regression line as:

$$V(R_i) = V(\overline{Y}) + (X_i - \overline{X})^2 V(b_i) + (X_i^2 - \overline{X}^2)^2 V(b_2), \tag{9}$$

The variance of G is given by the equation:

$$V(G) = 4\left\{V(b_1) + (X_1 + X_2)^2 V(b_2) + (X_1 + X_2)\, V(b_1 b_2)\right\}. \tag{10}$$

When the test is run using 1- and 2-ml doses of toxins, then $X_1 = 5$ and $X_2 = 6$. Under these conditions:

$$V(R_1) = 0.0134, \quad V(R_2) = 0.0018$$

and

$$V(G) = 0.0355$$

so that:

$$V(D) = \frac{1}{0.6352}\left\{V(N) + 0.0355D^2\right\} \tag{11}$$

and:

$$V(N) = 0.0134 + 0.0018 + V(T_1) + V(T_2). \tag{12}$$

Example. A sample of toxins of unknown potency was tested in this laboratory. It was known to kill Fischer rats in slightly more than 90 min when injected intravenously in doses of 1 ml at a concentration of 1X. The response of the unknown toxins was compared with the response curve described by equation 1. Each of three Fischer rats was injected with 1 ml of the test toxins, and their reciprocal response times in minutes were recorded (Figure 2). Three other Fischer rats were each

injected intravenously with 2 ml of the test toxins. Their reciprocal response times were also recorded (Figure 2). From these six reciprocal response times, values of T_1 and T_2 were calculated. Corresponding values of R_1 and R_2 were obtained from the regression line by substituting, respectively, the values 5 and 6 for X in equation 1. The value of L was calculated from equation 3 by use of the values 5 and 6 for X_1 and X_2. The values 5 and 6 were used in these two cases, because they are the $\log_2$ of the number of units in 1 and 2 ml of the reference toxins.

The value of D was calculated by substituting the previously calculated values of R_1, R_2, T_1, T_2, and L in equation 2. This value of D was found to be 0.78. This indicates that the test toxins were 0.78 $\log_2$ unit more potent than the reference. A 1-ml amount of the reference toxins contains 5 $\log_2$ units, so the test toxins must contain 5.78 $\log_2$ units. Thus, the test toxins have 55.0 potency units per ml at the concentration tested (5.78 X .301 = 1.73978 $\log_{10}$ units).

The formulas for calculating the variance of the estimate D of the $\log_2$ of the difference in potency between the test and the reference are described above as equations 6 through 10. These calculations were made in this example, and it was found that SE (D) = 0.26. Using normal theory, the 95% confidence limits of D become UL(D) = 1.30, and LL(D) = 0.26. From these the 95% confidence limits of P were calculated as UL(P) = 79.4 units per ml, and LL(P) = 38.0 units per ml.

DISCUSSION. Anthrax toxins are composed of at least three factors, I, II, and III, by the classification of Stanley and Smith (1961, 1963) or, respectively, edema factor, protective antigen, and lethal factor according to Beall et al. (1962). Both in vitro-produced toxins, as used in this report, and in vivo toxins, as reported by Klein et al. (1963), may be quantitated accurately. The procedure further provides an effective reference for quantitating natural resistance or relative immunity as described by Klein et al. (1963), because the absolute dose of toxins required to elicit a given response will bear a definite relationship to host resistance or susceptibility.

	Reference Toxin $Y = 100/t$ 1 ml.	2 ml.		Test Toxin $Y = 100/t$ 1 ml.	2 ml.
Rat 1			Rat 1	1.39	1.67
Rat 2			Rat 2	1.25	1.56
Rat 3			Rat 3	1.15	1.59
$\sum Y$			$\sum Y$	3.79	4.82
$\bar{Y} = R_1$	0.92	1.34	$\bar{Y} = T_1$	1.26	1.61
$R_1 + R_2 =$	2.26		$T_1 + T_2 =$	2.87	
$\sum Y^2$			$\sum Y^2$	4.8171	7.7506
$V(R_1)$	.0134	.0018	$V(T_1)$	.0048	.0011

$b_0 = -2.5912$
$b_1 = .9592$
$b_2 = -.0510$

$V(b_0) = .07712089$
$V(b_1) = .00951355$
$V(b_2) = .00006804$

$V(b_1 b_2) = -.000800$

$L = b_1 + b_2(x_1 + x_2)$

$x_1 =$ 5 $x_2 =$ 6 $(x_1 + x_2) =$ 11 $(x_1 + x_2)^2 =$ 121

$b_1 = 0.9592$
$b_2(x_1 + x_2) = 0.5607$
$L = 0.3985$
$2L = 0.7970$
$4L^2 = 0.6352$

$$D = \frac{(T_1 + T_2) - (R_1 + R_2)}{2L} = \frac{2.87 - 2.26}{0.7970} = 0.78$$

$D^2 = 0.6084$

$\text{Log}_2 P = 5 \pm D = 5$

$\text{Log}_{10} P = 0.301 \times 5.78 = 1.74$ $P = \boxed{55.0 \text{ U/ml}}$

$V(G) = 4\{V(b_1) + (x_1 + x_2)^2 V(b_2) + (x_1 + x_2) V(b_1 b_2)\} =$ 0.0335

$V(N) = V(R_1) + V(R_2) + V(T_1) + V(T_2) =$ 0.0211

$V(D) = \frac{1}{4L^2} \{V(N) + D^2 V(G)\} =$ 0.0672

$SE(D) =$ 0.26

$UL(D) =$ 1.30 $\text{Log}_{10} UL(P) =$ 1.90 $UL(P) =$ 79.4

$LL(D) =$ 0.26 $\text{Log}_{10} LL(P) =$ 1.58 $LL(P) =$ 38.0

FIG. 2. *Calculation form for potency of anthrax toxins.*

The biological activities of these compounds are numerous, and it is likely that some responses are still to be discovered. The problem of evaluating activity and mode of action of compounds which have a synergistic biological action is more difficult than for "single compounds." Quantitation, therefore, is important to allow the work of various investigators to be related more exactly to each other. The Fischer 344 rats are commercially available, and reference anthrax toxins will be provided for responsible investigators who desire to work with this material for use in establishing units. The methods used in this standardization of these toxins may be appropriate to the standardization of other biologically active toxins.

ACKNOWLEDGEMENTS. We thank Martha K. Ward, U. S. Public Health Service, and Dorothy M. Molnar for their cooperation and independent assay of the toxins, and Francis A. Beall who supplied his own animals and, in addition, independently assayed the toxins. Thanks are also extended to Bill G. Mahlandt and Charles C. Wigington for producing the anthrax toxins in vitro, and preparing them for storage.

LITERATURE CITED

Beall, F. A., M. J. Taylor, and C. B. Thorne. 1962. Rapid lethal effect in rats of a third component found upon fractionating the toxin of *Bacillus anthracis*. J. Bacteriol. 83:1274-1280.

Eckert, N. J., and P. F. Bonventre. 1963. *In vivo* effects of *Bacillus anthracis* culture filtrates. J. Infect. Dis. 103: 226-232.

Finney, D. J. 1952. Statistical method in biological assay. Hafner Publishing Co., New York.

Harris-Smith, P. W., H. Smith, and J. Keppie. 1958. Production *in vitro* of the toxin of *Bacillus anthracis* previously recognized *in vivo*. J. Gen. Microbiol. 19: 91-103.

Keppie, J., H. Smith, and P. W. Harris-Smith. 1955. The chemical basis of the virulence of *Bacillus anthracis*. III. The role of the terminal bacteremia in death of guinea pigs from anthrax. Brit. J. Exp. Pathol. 36: 315-322.

Klein, F., B. W. Haines, B. G. Mahlandt, I. A. DeArmon, Jr., and R. E. Lincoln. 1963. Dual nature of resistance mechanisms as revealed by studies of anthrax septicemia. J. Bacteril 85: 1032-1038.

Klein, F., D. R. Hodges, B. G. Mahlandt, W. I. Jones, B. W. Haines, and R. E. Lincoln. 1962. Anthrax toxin: causative agent in the death of rhesus monkeys. Science 138: 1331-1333.

Sargeant, K., J. L. Stanley, and H. Smith. 1960. The serological relationship between purified preparations of factor I and II of the anthrax toxin produced in vivo and in vitro. J. Gen. Microbiol. 22: 219-228.

Smith, H., and R. C. Gallop. 1956. The chemical basis of the virulence of Bacillus anthracis. VI. An extracellular immunizing agressin isolated from exudates. Brit. J. Exp. Pathol. 37: 144-155.

Smith H., J. Keppie, and J. L. Stanley. 1955a. The chemical basis of the virulence of Bacillus anthracis. V. The specific toxin produced by B. anthracis in vivo. Brit. J. Exp. Pathol. 36: 460-472.

Smith, H., J. Keppie, J. L. Stanley, and P. W. Harris-Smith. 1955b. The chemical basis of the virulence of Bacillus anthracis. IV. Secondary shock as a major factor in death of guinea pigs from anthrax. Brit. J. Exp. Pathol. 36: 323-255.

Smith, H., D. W. Tempest, J. L. Stanley, P. W. Harris-Smith, and R. C. Gallop. 1956. The chemical basis of the virulence of Bacillus anthracis. VII. Two components of the anthrax toxin: their relationship to known immunizing aggressins. Brit. J. Exp. Pathol. 37: 263-271.

Stanley, J. L., K. Sargeant, and H. Smith. 1960. Purification of factors I and II of the anthrax toxin produced in vivo. J. Gen. Microbiol. 22: 206-218.

Stanley, J. L., and H. Smith. 1961. Purification of factor I and recognition of a third factor of the anthrax toxin. J. Gen. Microbiol. 26: 49-66.

Stanley, J. L., and H. Smith. 1963. The three factors of anthrax toxin: their immunogenicity and lack of demonstratable enzymatic activity. J. Gen. Microbiol. 31: 329-337.

Taylor, M. J., G. G. Kennedy, and G. P. Blundell. 1961. Experimental anthrax in the rat. I. The rapid increase of natural resistance observed in young hosts. Amer. J. Pathol. 38: 469-480.

Thorne, C. B., D. M. Molnar, and R. E. Strange. 1960. Production of toxin _in vitro_ by _Bacillus anthracis_ and its separation into two components. J. Bacteriol. 79: 450-455.

AN INVESTIGATION OF THE DISTRIBUTION OF DIRECT HITS ON PERSONNEL BY SELF-DISPERSING BOMBLETS*

David M. Moss and Theodore W. Horner
Booz Allen Applied Research Inc.

ABSTRACT. The question has been raised concerning the lethal hazard to personnel from self-dispersing bomblets. The solution of this question involved the derivation of a distribution and the computation of parameters for a specific problem. The basic method used was to define a random variable, θ, the number of individuals which are hit:

$$\theta = \sum_{i=1}^{N} (1 - 0^{n_i}) ,$$

where N is total number of personnel and n_i is the number of bomblets striking the i^{th} individual. The moment-generating-function of this random variable was found and, hence, its distribution function. The distribution of casualties was found to be Poisson under the general assumptions of the problem.

* * * * * * * * * * * *

The question has been raised concerning the lethal hazard to personnel from self-dispersing bomblets by direct hits. In trying to determine the lethality of these bomblets many factors must be taken into account.

Among the factors which bear on this problem is that of protection. The flight of the bomblets might be intercepted by trees, buildings, or other natural or man-made obstructions, and would therefore descrease the chances of a lethal hit. In this study the interest is directed toward assessing the maximum hazard to personnel. It is, therefore, assumed that all personnel are completely exposed. It is also assumed that all personnel are in an upright position and no person provides any protection for another person. Thus, each person is completely and equally exposed to the possibility of a direct hit by a bomblet.

*Work on which paper is based was supported by contract with the U. S. Army Biological Laboratories, Fort Detrick, Frederick, Maryland.

Other assumptions made in order to assess the maximum hazard are that all personnel are within the target area of interest and all bomblets hit somewhere within this area. It can also be assumed that the vulnerable portions of an individual are his head and neck. If other portions of the body are struck, it is assumed that lethal damage is not inflicted.

The objective here will be to determine the hazard to personnel on target resulting from a drop of self-dispersing bomblets. The distribution of the number of lethal hits resulting from such a drop will be determined and in addition the expected number of such hits and the associated variance will be found. The results found will reflect the maximum hazard involved.

In addition to the theoretical work done here, the results for a specific case will be given. This will be the case where 600 bomblets are dropped on a one square kilometer area which contains 4000 persons.

First it will be assumed that there are N individuals in the target area, A_T. There are n bomblets dropped, all of which land in the target area. Further it will be assumed that bomblets and individuals are uniformly and independently distributed in the target area; however, it will be shown later that the individuals may assume any distribution. It will also be assumed that individuals and bomblets can be represented by circles with areas given by

(1) $$A_p = \pi r_1^2$$

and

(2) $$A_b = \pi r_2^2,$$

where r_1 is the radius of the critical area of an individual and these areas for all individuals are considered to be the same, and r_2 is the radius of a bomblet. Now in order to produce a casualty, the center of a bomblet must fall within the circle with radius

(3) $$r = r_1 + r_2 \quad .$$

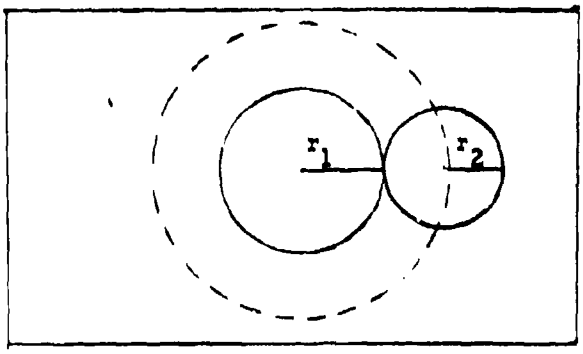


Figure 1
Casualty Radius Diagram

The target area can be divided up into N circular cells, each with radius $r = r_1 + r_2$ representing individuals, plus one cell which represents that part of the target in which there are no individuals. We can assign a value p_i to the probability that a bomblet falls in the i^{th} cell. Let n_i represent the number of bomblets that fall in the i^{th} cell. Then

$$(4) \qquad \sum_{i=1}^{N+1} p_i = 1$$

and

$$(5) \qquad \sum_{i=1}^{N+1} n_i = n ,$$

where n is the total number of bomblets.

The interest now is in the number of persons hit or the number of casualties, denoted here by θ. What is needed is a variable which will give the number of casualties, regardless of whether an individual is hit more than once. One such variable could be obtained by defining a variable which is either zero or one depending on whether an individual is missed or hit. If such a variable is then summed over all individuals, the result would be the total number of casualties, θ.

Note that

$$0^{n_i} = \begin{cases} 1 \text{ when } n_i = 0 \\ 0 \text{ when } n_i > 0 \end{cases} \tag{6}$$

and

$$1 - 0^{n_i} = \begin{cases} 0 \text{ when } n_i = 0 \\ 1 \text{ when } n_i > 0; \end{cases} \tag{7}$$

that is, if the number of hits of an individual is one or more, $(1-0^{n_i})$ will be one and will be zero otherwise. Thus, let us define our variable of interest as

$$\theta = \sum_{i=1}^{N} (1-0^{n_i}) . \tag{8}$$

This variable tells us the number of individuals which are hit and it is about this random variable that we want more information.

Now, before going on, let's look more closely at our probabilities, where p_i $(i = 1, 2, \ldots, N)$ defines the probability of a hit of the individual in the i^{th} cell. Obviously, the probability that any particular bomblet hits any particular individual is the same for all bomblets and all individuals. Also it is quite clear that the probability of a randomly chosen bomblet from a uniform distribution of bomblets hitting any individual is equal to the ratio of the area, A_c, of the circle with radius r to the total target area, A_T.

Thus

$$p = A_c/A_T , \tag{9}$$

where

(10) $$A_c = \pi(r_1 + r_2)^2 .$$

Note that since

(11) $$\sum_{i=1}^{N+1} p_i = 1$$

and the p_i's, $i = 1, 2, \ldots, N$, are equal, we thus have

(12) $$p_{N+1} = 1 - Np .$$

What we have is essentially the probability of a randomly selected point being within a certain area. Note in Figure 2 that the probability that a randomly selected point lies in a given circle is the same in A and B and also that the probability of at least x of the n points lying within a circle is the same in both. Based on this it can be seen that our results will be independent of the distribution of personnel.

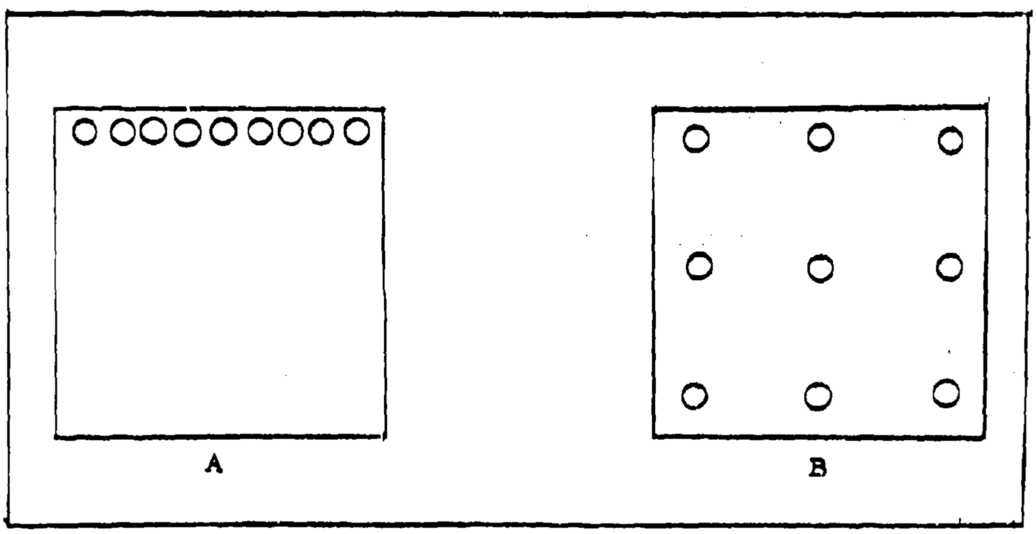


Figure 2
Possible Personnel Configurations

Now let us look at an analogous situation. Suppose that we have $N + 1$ cells into which we randomly throw n balls.

Table I
Distribution of Balls Falling into Cells

Cell	Probability	Number of Balls Falling Into Cell
1	$p_1 = p$	n_1
2	$p_2 = p$	n_2
.	.	.
.	.	.
.	.	.
N	$p_N = p$	n_N
N + 1	$p_{N+1} = 1 - Np$	$n_{N+1} = n - \sum_{i=1}^{N} n_i$

This is a multinomial situation where

$$f(n_1, n_2, \ldots, n_{N+1}) = \left(n! \Big/ \prod_{i=1}^{N+1} n_i! \right) p_1^{\,n_1} p_2^{\,n_2} \cdots p_{N+1}^{\,n_{N+1}} . \tag{13}$$

Since it is θ in which we are interested, we need to discover the distribution of θ. The approach taken here will be to find the moment-generating-function of θ and from it the distribution of θ.

Recalling the definition of moment-generating-function from mathematical statistics and substituting for θ from equation (8), we have

$$
\begin{aligned}
M_\theta(t) &= E\left\{e^{t\theta}\right\} \\
&= E\left\{\exp\left[t\sum_{i=1}^{N}(1-0^{n_i})\right]\right\} \\
\text{(14)}\qquad &= e^{tN}E\left\{\exp\left[-t\sum_{i=1}^{N}0^{n_i}\right]\right\} \\
&= e^{tN}E\left\{\prod_{i=1}^{N}e^{-t0^{n_i}}\right\}.
\end{aligned}
$$

Now

$$
\text{(15)}\qquad e^{-t0^{n_i}} = \begin{cases} e^{-t} & \text{when } n_i = 0 \\ 1 & \text{when } n_i > 0, \end{cases}
$$

and equivalently

$$
\text{(16)}\qquad e^{-t0^{n_i}} = 1 + 0^{n_i}(e^{-t}-1).
$$

Note that (16) holds identically and that the right hand side is not part of a series expansion. Substituting back in (14), we have

(17) $$M_\theta(t) = e^{tN} E \left\{ \prod_{i=1}^{N} \left[1 + 0^{n_i}(e^{-t}-1) \right] \right\} .$$

Now let

(18) $$b_i = 0^{n_i}(e^{-t}-1)$$

and substitute in (17):

$$M_\theta(t) = e^{tN} E \left\{ \prod_{i=1}^{N} (1 + b_i) \right\}$$

(19) $$= e^{tN} E \left\{ 1 + \sum_{i=1}^{N} b_i + \sum_{\substack{i \\ j>}} \sum_{j} b_i b_j + \sum_{i} \sum_{j} \sum_{\substack{k \\ k>j>i}} b_i b_j b_k + \ldots + \sum_{i} \sum_{j} \ldots \sum_{\substack{m \\ m> \ldots >j>i}} b_i b_j \ldots b_m \right\} .$$

Now taking the expectation of a typical term, say the $g + 1^{st}$ term and substituting from (18), we have

$$T_{g+1} = E\left\{\underbrace{\Sigma\,\Sigma\,\ldots\,\Sigma}_{g}\ b_i b_j \ldots b_m\right\}$$

(20) $$= \Sigma\,\Sigma\,\ldots\,\Sigma\ E\{b_i b_j \ldots b_m\}$$

$$= \Sigma\,\Sigma\,\ldots\,\Sigma\ E\left\{(e^{-t}-1)^g 0^{n_1} 0^{n_2} \ldots 0^{n_g}\right\}$$

$$= (e^{-t}-1)^g\, \Sigma\,\Sigma\,\ldots\,\Sigma\ E\left\{0^{n_1} 0^{n_2} \ldots 0^{n_g}\right\} .$$

Now the expectation of the last factor in (20) is

$$E\,\{0^{n_1} 0^{n_2} \ldots 0^{n_g}\} = \Sigma\ 0^{n_1} 0^{n_2} \ldots 0^{n_g} f(n_1, n_2, \ldots, n_{N+1})$$

which becomes upon substitution from (13)

$$E\left\{0^{n_1} 0^{n_2} \ldots 0^{n_g}\right\} = \Sigma\ 0^{n_1} 0^{n_2} \ldots 0^{n_g} \frac{n!}{\Pi n_i!}\ p_1^{n_1} p_2^{n_2} \cdots p_{N+1}^{n_{N+1}}$$

(21) $$= \Sigma \frac{n!}{\Pi n_i!} (0 \cdot p_1)^{n_1} (0\, p_2)^{n_2} \ldots (0 \cdot p_g)^{n_g} \cdot p_{g+1}^{n_{g+1}} \cdots \cdots p_{N+1}^{n_{N+1}}$$

$$= \left(0 + 0 + \ldots + 0 + p_{g+1} + \cdots + p_N + p_{N+1}\right)^n$$

$$= (1 - gp)^n ,$$

the last step following from the fact that

$$\sum_{i=1}^{N+1} p_i = 1 . \tag{22}$$

Substituting the result from (21) back in (20) we get

$$\begin{aligned} T_{g+1} &= (e^{-t}-1)^g \Sigma\Sigma \ldots \Sigma (1 - gp)^n \\ &= (e^{-t}-1)^g (1 - gp)^n \Sigma\Sigma \ldots \Sigma (1) \\ &= (e^{-t}-1)^g (1 - gp)^n \binom{N}{g} . \end{aligned} \tag{23}$$

Using this last result in equation (19), we now find the moment-generating-function to be

$$\begin{aligned} M_\theta(t) &= e^{tN} \left\{ 1 + N(e^{-t}-1)(1-p) \right. \\ &\quad + \binom{N}{2} (e^{-t}-1)^2 (1-2p)^n + \binom{N}{3} (e^{-t}-1)^3 (1-3p)^n \\ &\quad \left. + \ldots + (e^{-t}-1)^N (1-Np)^n \right\} \\ &= e^{tN} \sum_{g=0}^{N} \binom{N}{g} (e^{-t}-1)^g (1 - gp)^n . \end{aligned} \tag{24}$$

The maximum value of gp is Np. However, Np is extremely small as seen from the example following the theory. Since, therefore, gp is extremely small,

(25) $$(1 - gp)^n \simeq e^{-npg};$$

which follows because

$$e^{-ngp} = (e^{-gp})^n$$

(26) $$= (1 - gp + \frac{(gp)^2}{2!} - \frac{(gp)^3}{3!} + \dots)^n$$

$$\simeq (1 - gp)^n.$$

Therefore

$$M_\theta(t) \simeq e^{tN} \sum_{g=0}^{N} \binom{N}{g} (e^{-t}-1)^g (e^{-np})^g (1)^{N-g}$$

$$= e^{tN} \sum_{g=0}^{N} \binom{N}{g} \left[e^{-np}(e^{-t}-1)\right]^g (1)^{N-g}$$

(27) $$= e^{tN} \left[e^{-np}(e^{-t}-1) + 1\right]^N$$

$$= \left[e^t e^{-np}(e^{-t}-1) + e^t\right]^N$$

$$= \left[e^{-np}(1-e^t) + e^t\right]^N$$

$$= \left[e^{-np} - e^t e^{-np} + e^t\right]^N$$

$$= \left[e^{-np} + e^t(1 - e^{-np})\right]^N.$$

Now in the above result let

(28) $$Q = e^{-np}$$
and
$$P = 1 - e^{-np}$$

We then have

(29) $$M_\theta(t) \simeq (Q + Pe^t)^N$$

which can be recognized as the moment-generating-function for the binomial distribution. Thus θ is approximately binomially distributed with parameters P, Q, and N. The expected value of θ or the mean number of casualties is given by

$$E\{\theta\} = NP$$

(30) $$= N(1 - e^{-np})$$

$$= N\left[1 - \left(1 - np + \frac{(np)^2}{2!} - \frac{(np)^3}{3!} + \ldots\right)\right]$$

$$\cong Nnp,$$

the last step following since np is extremely small. Thus the $E(\theta)$ is small unless N is extremely large. Also because P is small, the distribution of θ can be approximated by a Poisson distribution and therefore the variance is also approximately Nnp. The distribution of θ, where θ is the number of casualties, is given by

(31) $$p(\theta) = (Nnp)^\theta \; e^{-Nnp}/\theta!$$

Now let's look at the specific problem: namely that of dropping 600 bomblets on a one square kilometer target which contains 4000 personnel. It is given that:

a. $A_T = 10^{10}$ cm^2

b. $A_p = 314$ cm^2

c. $N = 4 \times 10^3$

d. $n = 6 \times 10^2$

e. $r_1 = 10$ cm

f. $r_2 \approx 7$ cm.

From these it is found that

$$\begin{aligned} p &= A_c/A_T \\ &= \pi\,(10 + 7)^2 \;/\; 10^{10} \\ &= 9.1 \times 10^{-8}\,, \end{aligned}$$

and that

$$\begin{aligned} E\,\{\theta\} &= Nnp \\ &= (4 \times 10^3)\,(6 \times 10^2)\,(9.1 \times 10^{-8}) \\ &= 0.22 \end{aligned}$$

and

$$VAR\,\{\theta\} = 0.22.$$

Note that N_p, which is the maximum value of gp, is $Np = 3.64 \times 10^{-4}$, a very small quantity. Further, it is found that the probability of exactly x casualties under the given assumptions are as in Table II.

Table II
Casualty Distribution

Number of Casualties	Probability of Occurrence
0	0.80252
1	0.17655
2	0.01942
3	0.00142
4	0.00008
5	0.00000

Note that the expected number of casualties, 0.22, is approximately 0.0055 percent of the 4000 personnel or approximately one casualty in five similar drops.

The hazard to personnel resulting from a drop of self-dispersing bomblets was found to be very low. It was found that the number of casualties, θ, is Poisson distributed of form

$$p(\theta) = (Nnp)^{\theta} e^{-Nnp}/\theta ! ,$$

where N is the number of personnel on target, n is the total number of bomblets, and p is the probability that an individual is hit by a particular bomblet. For the specific case of 600 bomblets and 4000 persons in a one square kilometer area, p is approximately 9.1×10^{-8} and the expected number of casualties is 0.22.

EXPLOSIVE SAFETY AND RELIABILITY ESTIMATES FROM A LIMITED SIZE SAMPLE

J. N. Ayres, L. D. Hampton, I. Kabik
U. S. Naval Ordnance Laboratory
White Oak, Silver Spring, Maryland

ABSTRACT. The problem of predicting, from small sample testing, high reliability and/or high safety for explosive items is becoming more acute. Often the available test sample is no greater than 200. Only a single test per item is allowable and the data is always of the go/no-go variety. Methods being used for making conservative extrapolations to the high and low probability of firing points are reviewed and illustrated. The question of how to do the job better is posed and left to the clinicians for answer.

INTRODUCTION. The problem which we wish to present is how to make, with small samples, reasonable estimates of the stimuli corresponding to the high and low probability of firing of electro-explosive devices (EED's).

A typical EED is shown in Fig. 1. Essentially, it consists of an insulator carrying two electrical conductors across which is attached a resistance wire. Surrounding the resistance wire is a sensitive explosive. When electrical energy is dissipated in the wire, the resultant temperature rise causes the explosive to heat and react chemically, and thus produce an explosion.

EED's are used by the military for a number of purposes; to cause detonation of explosive loaded shells, bombs, grenades, missiles, mines, etc., to ignite propellants for guns and rockets, to close switches such as in fuze arming circuits, to release stores from aircraft, to eject pilots from aircraft, and to separate missile stages. These are only some of the more common uses.

The designer of explosive ordnance has always been faced with the problem of estimating the safety and reliability of his explosive system. The safety and reliability associated with the EED of electrically operated explosive ordnance, are, of course, important links in this system. For reasons to be given, estimating the safety and reliability to be expected from an EED subjected to various stimuli is usually not simple. The ordnance designer in the past has often overcome lack of information on

reliability at least, by the numbers of items strategically used, i.e., the number of shells fired or the number of bombs dropped, etc. Thus unreliability could be compensated for in actual field usage.

Modern weapons and warfare, however, have introduced new problems. It is too costly to fire large numbers of expensive ordnance devices: the catastrophic results of a safety failure of certain types of munitions are intolerable; the intensity of certain stimulis which may cause inadvertent firing (electro-magnetic radiation from radars for example) has increased tremendously in the last decade and is slated to increase further. These changes have made it virtually mandatory that reasonable estimates of response of EED's to electrical stimuli be made.

RELEVANT FACTS.

(a) For economic reasons it is impossible to make a direct demonstration of the response of interest. The stimulus for reliability of 99.9+% is usually desired at 95% confidence. Conversely, safety may demand estimates at 95% confidence of the stimulus at which no more than 1 in a million devices would be expected to fire. Funds are never available to run direct demonstration tests.

(b) The nature of EED's preclude repeated testing on a single device. Since these systems respond chemically to temperature elevation at the resistance wires, it is not known, once a single test at a given stimulus was large enough to have altered the EED's response characteristics. It must therefore be assumed that the possibility of alteration is great enough to preclude more than one test on a given EED. The only piece of information thus possible from each single test is either the EED fired or failed at that particular test stimulus.

(c) It has been found[1], from a large number of firings on EED's (approx. 10,000 firings of Squib Mk I), that no standard distribution function fits exactly the tails of the observed EED stimulus-response distribution. A number of distribution functions have been tested for their conformance to the experimental firing data. They all fail at the tails of the curve, see Fig. 2. But it is precisely these regions of the distribution which we must estimate.

(d) Usually no more than 200 test samples are available to make estimates on one side of the mean firing (50%) point, whether high or low. Even a sample size of 200 is sometimes very difficult to obtain and may be quite expensive.

(e) Popular test schemes, such as the "Bruceton" test[2], which are conservative of sample size, often give poor estimates because of long extrapolation, poor estimate of the standard deviation, and/or non-applicability of the selected underlying distribution[3,4].

THE PROBLEM. By now it should be obvious that we must make multi-million dollar estimates on tens or hundreds of dollars worth of data. We must design our experiments so that we most wisely expend our available samples so that we can minimize the error of making extrapolations to the desired answer. We realize that extrapolation is at best a risky business but: is there any other choice?

In the following section we will tell you what we think we know and the methods we are now using.

The basic problem is to collect data which will permit the computation of the variation of the probability of firing as a function of the firing stimulus. It is desirable to allocate the samples so that the data collected will be as close as possible to the functioning level(s) we wish to estimate. Ideally we should collect go/no-go data at a number of stimulus levels. As shown in Fig. 3, we wish to estimate the stimulus, Xe, at which we can expect a high level of response, Ye. We show data collected at five levels of stimulus X_1, $X_2 \ldots X_5$. A line has been fit to the observed data and at point Xe, Ye on this line, is the intersection which gives us the desired stimulus value.

The process of drawing the straight line shown in Fig. 3, and making the indicated prediction implicitly makes the following assumptions.

1. That there is no sampling error,
2. That the distribution function is chosen correctly, and
3. That there is no systematic error in the instrumentation or testing procedure.

But we know that there must be some sort of error simply because the data points do not fall on the line. By performing the "Chi-Square" statistical test on the data we can decide whether or not the observed variability (scatter) is what might be expected from sampling error alone. If this is the case, then we can draw an appropriate confidence band as in Fig. 3.

WHAT WE HAVE DONE. But rather than multi-point testing we have made what we believe to be conservative estimates of extreme probability of firing points by the test and extrapolation procedures given below.

To minimize the importance of assumptions regarding the frequency distribution it is again desirable to base these estimates on data taken as close as possible to the per cent point to be determined. The simplest such test would be one which calls for testing at two stimulus levels near the region in question. One of the two levels will be farther from the mean and closer to the desired point than the other. This will be designated the remote stimulus level. The data obtained can then be extrapolated to determine the stimulus associated with the desired per cent point. In planning such an experiment the following conditions should be met:

a. The difference between the stimuli used should not be small compared to the extrapolation distance (the difference between the desired point and the observed remote stimulus).

b. The number of trials at the remote stimulus level and the expected response at this level should be chosen so that the probability of observing a saturated level (either all-fires or all fails) is small*.

c. The number of trials made at the remote functioning level should be greater than the number of trials at the level closer to the mean in an attempt to obtain equal weighting of the two levels. A good choice is to take the number so that the product np (1-p) is the same for both levels, where n is the number of trials and p is the expected probability of fire.

---

*If a saturated level is observed, one trial can be converted to 1/2 fire + 1/2 fail. Or another, reversed, trial can be arbitrarily added to the data. Either method will give a conservative result.

It is assumed that only two hundred samples are available to estimate either an extremely high or else an extremely low probability of firing. The general procedure will be illustrated below for a high probability point; a numerical example is given in Appendix A.

a. Run a preliminary Bruceton type test on 20 samples using a log transform for the dosage*.

b. Use the Bruceton results to estimate the $\bar{X}+0.2s$, $\bar{X}+0.4s$, and $\bar{X}+1.3s$ levels**.

c. Test 50 EED's at the computed $\bar{X}+0.4s$ level.

d. If more than 5 fail, test 130 samples at the above calculated $\bar{X}+1.3s$ level.

e. If 5 or fewer failures occur, continue testing until 130 samples have been tested, and test 50 at the calculated $\bar{X}+0.2s$ level.

f. Using a log-logistic[5] probability space, plot the two points.

g. Extrapolate the straight line through the points so obtained to the desired probability or stimulus value.

By using only two points we have no way of applying the chi-square test. Nor can we draw the confidence band without a further assumption. To obtain more conservatism, two methods have been used.

Heterogeneity Assumption

We proceed as above but assume a heterogeneity factor*** of 1 in the equation for the confidence limit. This assumption allows computation and drawing of the confidence band as in Fig. 4. Implicit

---

*Considerable testing has led us to believe a logarithmic dosage to stimulus transform is of proper form.

**For low probability estimates these terms would be $\bar{X}-0.2s$, $\bar{X}-0.4s$, $\bar{X}-1.3s$, and following computations would be consistent.

*** $F = \frac{\chi^2}{n-2}$, where F = heterogeneity factor and n = the number of test levels.

in the assumption are the assumptions previously given also, i. e., we have chosen the correct distribution function; there is no systematic error in the instrumentation and test procedure; and only normal sampling error occurs.

Binomial Method

Using the second method of gaining conservatism, rather than plotting the measured points directly, calculate, at a desired confidence level (say 75%), the one-sided lower value of the higher percentage firing point, and the one-sided upper value of the lower percent firing point. Plot these points in a log-logistic probability space. Draw the straight line through these points and extrapolate to the desired value. See Fig. 5.

It is, of course, possible that if too conservative a value be set for the confidence limits of the upper one-sided, lower and the lower one-sided, higher per cent firing points, the slope of the line drawn through these limits will be negative. Such a situation, when it occurs, is not realistic and this more conservative estimating technique should be abandoned.

Our experience has shown us that although the logistic distribution function does not give an accurate fit to EED distribution functions at the tails, it at least errors on the conservative side, i. e., it will predict a lower safety than actually exists and a lower reliability than actually exists.

The two-level test and analysis, then, is one technique which we have used to make, with limited samples, estimates of extreme probability of firing points. We could certainly devise more elaborate and sophisticated variations, but we wonder if those more skilled than we in statistical theory might not be able to recommend alternate procedures which can do the job better. More specifically, we have wondered about, and have planned to work on, the application of non-parametric statistical methods to the problem. The clinic's opinion and advice on this matter could be beneficial since, at the time of this writing (June 1964), we are only in the preliminary thinking stage.

Finally, we have been hopeful that some combination might be made of statistics and the underlying physics of the mechanism by which wire bridge EED's function, to put bounds on the degree of extrapolation

needed in making our estimates. In this regard our work has shown that the heating of a wire bridge EED can be represented by the mathematical equation:

$$C_p \frac{d\theta}{dt} + \gamma\theta = p(t)$$

where C_p = heat capacity of bridge plug explosive

θ = temperature elevation above ambient

t = time

γ = heat loss factor, and $p(t)$ = power input.

The combination of this equation[6, 7] with Bowden's hot spot theory of explosions[8] has led to fairly accurate representation of EED firing characteristics over a limited range of input times (i. e., average powers). Since equipment is available for making independent measurements of C_p, γ, and C_p/γ, the cooling time constant, it appears possible to measure, on individual EED's, parameters which should be directly related to their individual firing characteristics.

REFERENCES

1. L. D. Hampton, J. N. Ayres, "Characterization of Squib Mk 1 Mod 0: Determination of the Statistical Model", NavWeps Report 7347, 30 January 1961.

2. Statistical Research Group, Princeton University, "Statistical Analysis for a New Procedure in Sensitivity Experiments", AMP Report 101.1R, SRG-P No. 40, (OSRD Report 4040) July 1944.

3. J. N. Ayres, L. D. Hampton, I. Kabik, "The Prediction of Very Low EED Firing Probabilities", NOLTR 63-133, 4 September 1963.

4. L. D. Hampton, J. N. Ayres, I. Kabik, "Estimation of High and Low Probability EED Functioning Levels", NOLTR 63-266, 3 February 1964.

5. Jour. of American Statistical Association, Vol. 48, p. 565-599, 1953, Joseph Berkson, "A Statistically Precise and Relatively Simple Method of Estimating the Bioassay with Quantal Response Based on the Logistic Function".

6. I. Kabik, L. A. Rosenthal, A. D. Solem, "The Response of Electro-Explosive Devices to Transient Electrical Pulses", 4th Navy Science Symposium, Pasadena, Calif., March 1960.

7. I. Kabik, L. Rosenthal, A. Solem, "The Response of Electro-Explosive Devices to Transient Electrical Pulses", NOLTR 61-20, 17 April 1961.

8. <u>Initiation and Growth of Explosion</u>, Bowden & Yoffe, Cambridge at the University Press, 1952.

APPENDIX A

ILLUSTRATIVE EXAMPLE

The units used for X are in terms of the transformed variable.

The twenty trial Bruceton gave a mean of 20.314 and standard deviation of 0.589.

The two test levels are then

$m+0.4s = 20.55$
$m+1.3s = 21.08.$

The results at these levels were

Near level 35/50 = 70%
Remote level 113/130 = 86.92%.

The upper 95% confidence limit at the near level is 78.68%. The lower 95% confidence limit at the remote level is 81.94%.

A straight line through the observed points is

$$Y = 1.13019\,X - 22.7014$$

(Y in Normits).

This gives estimates as follows:

95% point	21.542
99% point	22.144
99.99% point	23.347.

The equation for the lower 95% confidence band assuming the heterogeneity factor to be unity is

$$Y = 1.13019\,X - 22.7014 - 1.645\sqrt{0.014909 + 0.213434(X - 20.857)^2}$$

This gives estimates as follows:

95% point	22.9
99% point	24.8
99.99% point	29.0 .

The straight line through the binomial limits on the observed points has the equation

$$Y = 0.2226\,X - 3.7784\ .$$

This gives the following estimates

95% point	24.37
99% point	27.43
99.99% point	33.69 .

Using the same data with the logistic assumption, we have the following analysis

at the near level 35/50 = 70%

$$L = \ln \frac{35}{15} = 0.8473$$

at the remote level 113/130 = 86.92%

$$L = \ln \frac{113}{17} = 1.8942\ .$$

The straight line through these points is

$$L = 1.975\,X - 39.7447\ .$$

This gives the following estimates

	L	X
95%	2.9444	21.6
99%	4.5951	22.4
99.99%	9.2102	24.8 .

The binomial confidence limits as before are

near level 1.306; remote level 1.512 .

The straight line through these points is

$$L = 0.389\,X - 6.688$$

which gives

95% point	24.8
99% point	29.0
99.99% point	40.9 .

The hyperbola for the lower 95% confidence band has the equation

$$L = 1.9753\,X - 39.7447 - 1.645\sqrt{0.039557+0.579926(X-21.86)^2}$$

which has an asymptote

$$L = 0.723\,X - 13.6224 .$$

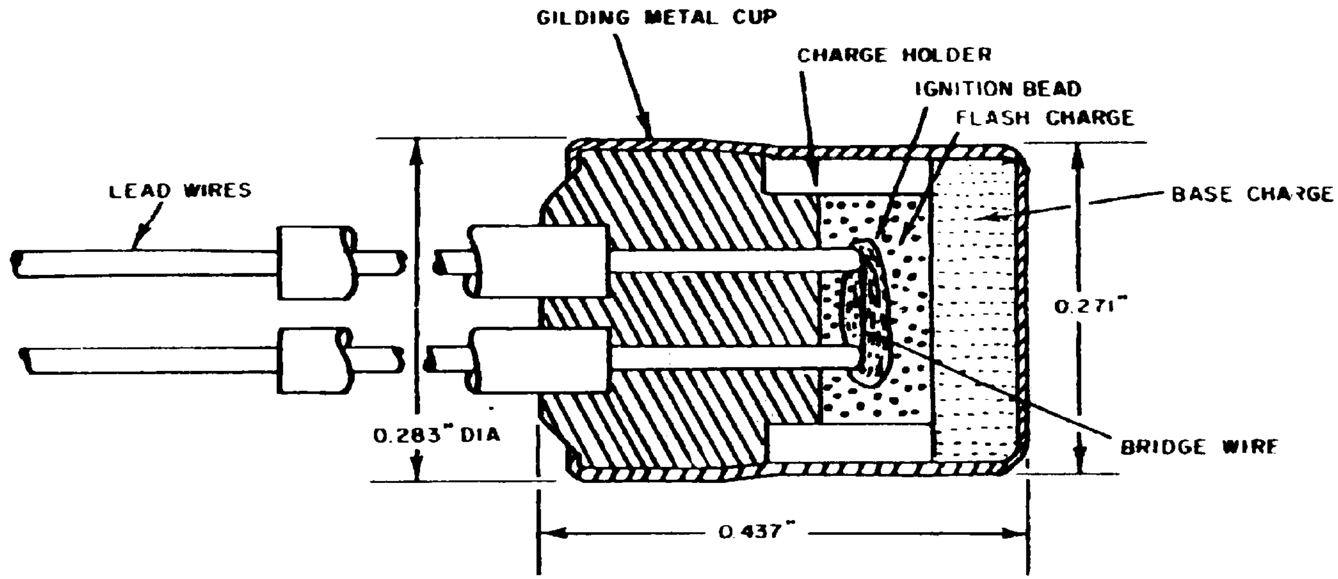
Estimates are

95% point	22.9
99% point	25.2
99.99% point	31.6

Summary of these calculations results

	Normal			Logistic		
	Straight Line	95% conf. band	Binomial	Straight line	Conf. Band	Binomial
95% point	21.54	22.9	24.4	21.6	22.9	24.8
99% point	22.14	24.8	27.4	22.4	25.2	29.0
99.99% point	23.35	29.0	33.7	24.8	31.6	40.9

Comparison of these values shows the more conservative nature of the logistic distribution. The difference is not marked at the 95% point but does show up at the more extreme points.



NOTES:

1. IGNITION BEAD - APPROX. 5 MG DDNP/$KClO_3$
2. FLASH CHARGE - APPROX. 45 MG BLACK POWDER
3. BASE CHARGE - APPROX. 45 MG BLACK POWDER
4. BRIDGE WIRE - 0.001" PLATINUM-IRIDIUM 0.060" LONG

FIG. I SQUIB MK I MOD O

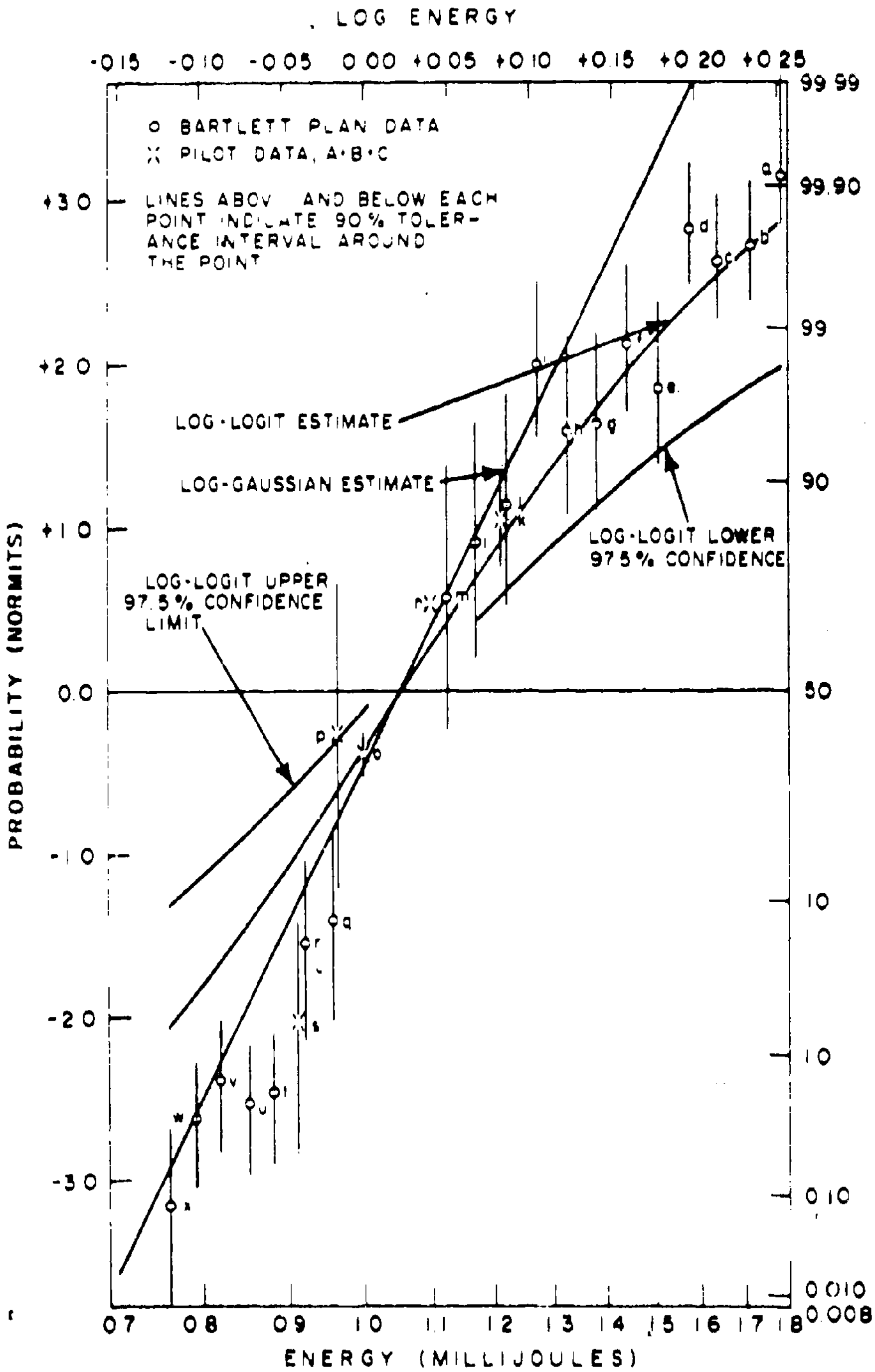


FIG. 2 LOG-GAUSSIAN, LOG-LOGISTIC FITS OF FIRING DATA

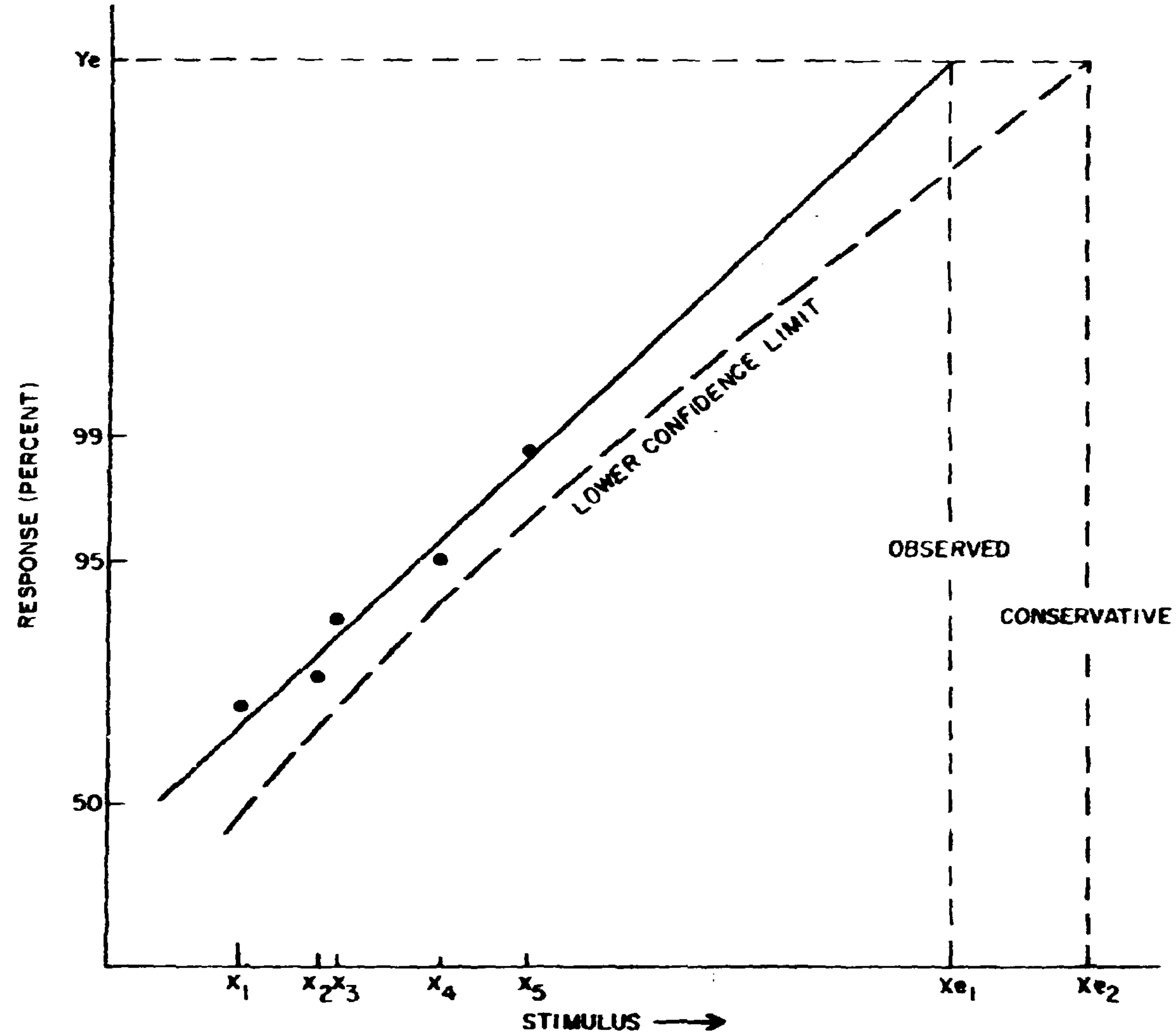


FIG. 3 MULTIPOINT - DATA ESTIMATE

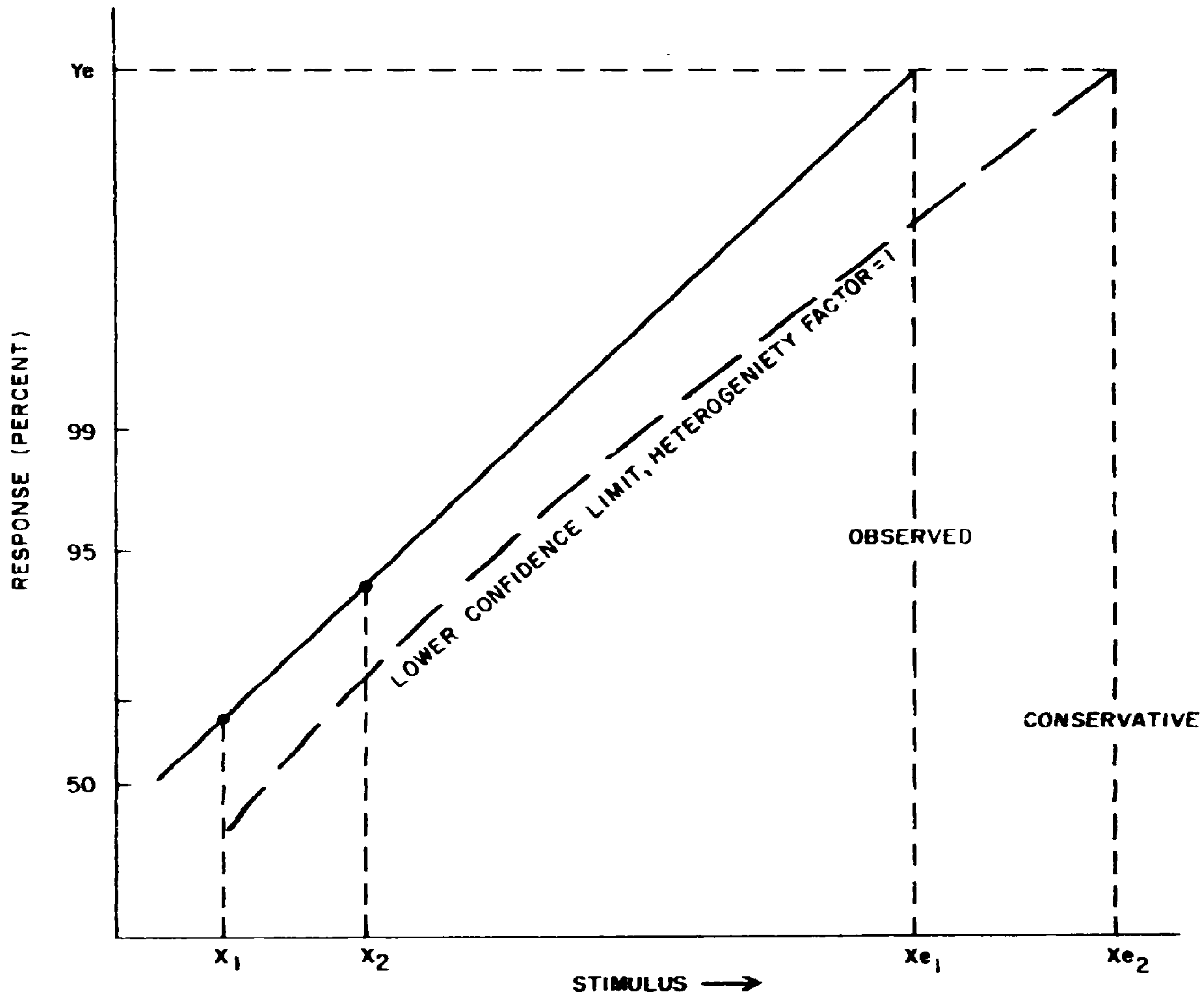


FIG. 4 2-POINT ESTIMATE WITH CONFIDENCE LIMIT

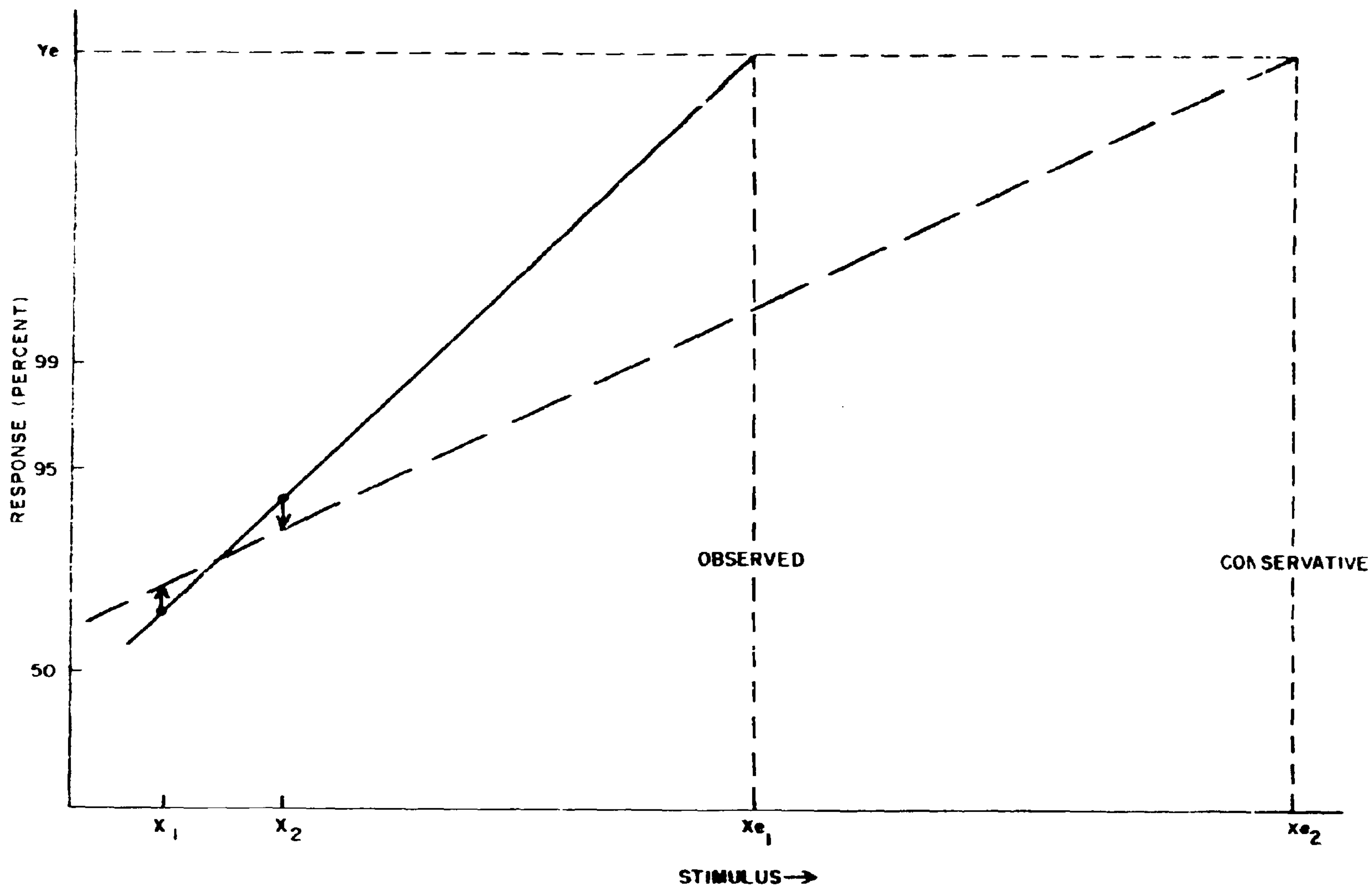


FIG. 5 2-POINT ESTIMATE WITH BINOMIAL LIMITS

CYCLIC DESIGNS*

H. A. David and F. W. Wolock
University of North Carolina at Chapel Hill and Boston College

1. INTRODUCTION. Cyclic designs are incomplete block designs consisting in the simplest case of a set of blocks obtained by cyclic development of an initial block. More generally, a cyclic design consists of combinations of such sets and will be said to be of size (n, k, r), where n is the number of treatments, k the block size, and r the number of replications.

It is well known (e.g. Bose and Nair [2]) that cyclic development of a suitably chosen initial block is one of the methods of generating designs with a high degree of balance in the arrangement of the treatments such as balanced incomplete block (BIB) designs and partially balanced incomplete block designs with two associate classes (PBIB(2) designs). Again, the cyclic type is a rather junior partner among the five types into which Bose and Shimamoto [3] classify PBIB (2) designs. The emphasis in these and many related papers has been understandably on the number of associate classes, the cyclic aspect being incidental. In the present article we proceed in opposite fashion putting the cyclic property first. It will be shown how cyclic designs may be systematically generated and how the non-isomorphic designs of given size may be enumerated and constructed. All such designs are PBIB designs but may have up to $\frac{1}{2}n$ associate classes. For $n \leq 15$ and $k = 3, 4, 5$, tables of the most efficient cyclic designs are presented and comparisons with BIB and PBIB (2) designs are made.

Points which make cyclic designs attractive are:

(i) Flexibility. A cyclic design of size (n, k, ik) exists for all positive integers n, k, i. If n and k have a common divisor d then a "fractional set" of size $(n, k, k/d)$ exists corresponding to each d. Fractional sets may be combined with designs of size (n, k, ik) to form fresh designs, or used by themselves especially if n is large. Thus there are cyclic designs for many sizes (n, k, r) for which no PBIB (2) design is available, but the reverse may also happen.

---

*Research supported by the Army Research Office, Durham, and the National Institutes of Health. This paper has been submitted for publication in the "Annals of Mathematical Statistics."

(ii) Ease of representation. No plan of the experimental layout is needed since the initial block or blocks suffice.

(iii) Youden type. In view of their method of generation cyclic sets with $r = k$, and hence combinations of such sets, provide automatic elimination of heterogeneity in two directions.

(iv) Analysis. For cyclic designs the coefficient matrix of the normal equations is a circulix. The inverse matrix may therefore be obtained explicitly (as another circulix), thus making possible a general method of analysis. Questions of analysis will not be considered further here since methods given in a special case by Kempthorne [9] continue to apply with minor modifications. However, details and aids to analysis are presented in [12].

Cyclic designs as a class in their own right were introduced for $k = 2$ by Kempthorne [9] and Zoellner and Kempthorne [13]. Design aspects for the case $k = 2$, which has some special features, were considered in [6] and [7], and will not be treated in this paper. For general k cyclic designs are closely related to the circular designs of Das [5]. See also the survey of non-orthogonal designs by Pearce [11] who calls cyclic designs a "little publicized class." PBIB designs have been studied from an algebraic point of view in a series of papers by Masuyama. In some of these (e.g. [10]) reference is made to cyclic designs but no detailed results are obtained.

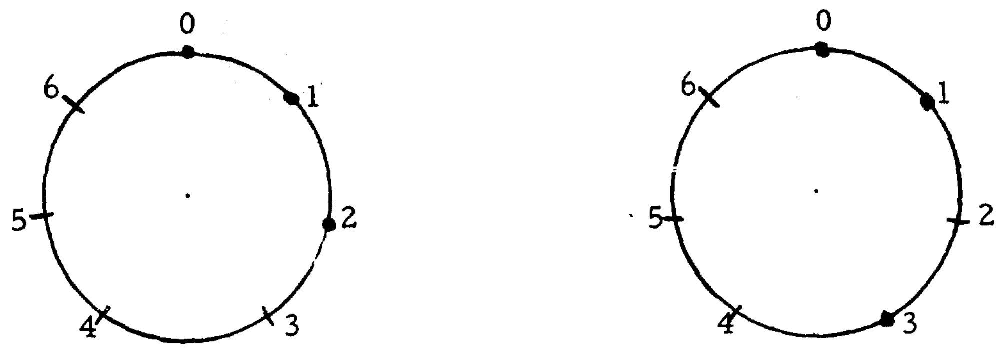
2. CYCLIC SETS. Label the treatments 0, 1, 2, ..., n-1. To fix ideas consider the arrangement of $n = 7$ treatments in blocks of size $k = 3$. The complete design of $\binom{7}{3} = 35$ distinct blocks may be set out as follows:

	{012}	:	012	123	234	345	456	560	601
	{013}	:	013	124	235	346	450	561	602
(1)	{014}	:	014	125	236	340	451	562	603
	{015}	:	015	126	230	341	452	563	604
	{024}	:	024	135	246	350	461	502	613

From any block the others in the same row may be obtained by increasing each object label in turn by 1, 2, 3, 4, 5, 6, and reducing modulo 7. The

rows have been arranged to start with the block of lowest numerical value and are designated by the initial block placed in braces. We call each row a <u>cyclic set</u>.

A block may also be conveniently represented by identical beads spaced regularly on a circular necklace. Fig. 1 shows the blocks 012 and 123.



012 = (115) 013 = (124)

Figure 1

The set {012} is then generated by successive unit rotations.

It is not difficult to show that each cycle set forms a partially balanced incomplete block (PBIB) design with b (no. of blocks) = n and r (no. of replications) = k. If objects i and j are α-th associates so are i and n-j. Thus the number m of associate classes is at most $\frac{1}{2}(n-1)$ for n odd and $\frac{1}{2}n$ for n even, but may be less, with m = 1 for a balanced (BIB) design. An additional feature of a cyclic set is that each object occurs once in each position within a block. Order effects are therefore automatically balanced out and the sets are Youden Type designs, balanced (m = 1) or partially balanced (m > 1).

The same procedure can be used for any n and k except that when n and k are not relative primes fractional sets arise consisting of n/d blocks, where d is any common divisor of n and k. In terms of Fig. 1 such sets correspond to arrangements of beads which can be reproduced in fewer than n rotations of the necklace.

For the purpose of systematically enumerating all cyclic sets it is convenient to characterize each set by a circular partition of n. Thus we may replace $\{0x_1\ x_2\ x_3 \ldots x_{k-2}\ x_{k-1}\}$ by $(x_1,\ x_2-x_1,\ x_3-x_2, \ldots,$ $x_{k-1}-x_{k-2},\ n-x_{k-1})$.

Example 1. For n = 8, k = 4 the set {0123} becomes (1115). The cyclic sets may now be written down in increasing order of the numerical value of the corresponding partition: (1115), (1124), (1133), (1142), (1214), (1223), (1232), (1313), (1322), (2222). After (1142) we omit (1151) this being identical with (1115), etc. As the repetition of digits indicates the set (1313) consists of the 4 blocks

0145 1256 2367 3470 (r = 2)

and (2222) of the 2 (disconnected) blocks 0246, 1357 (r = 1). These are still PBIB designs but, of course, no longer of the Youden Type. We shall say that the corresponding arrangements of beads on a necklace have periods 4 and 2, respectively. As a check note that all $\binom{8}{4}$ blocks are accounted for since 8x8+4+2 = 70.

For any n and k, the total number of sets, being equal to the number of distinct arrangements of k white beads and n-k black beads on a necklace of n beads (which may not be turned over) is given by (Jablonski [8])

$$N(k,\ n-k) = \frac{1}{n} \sum \phi(d) \frac{(n/d)!}{(k/d)!\ [(n-k)/d]!}, \tag{2}$$

where the summation is over all integers d (including unity) which are divisors of both k and n-k, and $\phi(x)$ is Euler's function, the number of integers less than and prime to x. Thus

$$N(4,\ 4) = \frac{1}{8}\left(\frac{8!}{4!\ 4!} + \frac{4!}{2!\ 2!} + 2.\ \frac{2!}{1!\ 1!}\right) = 10.$$

The number of cyclic sets of various sizes making up this total is tabulated in [7] for n ≤ 15.

If a design of size $n = b = 7$ and $k = r = 3$ is required a look at the association schemes of the 5 sets in (1) leads to $\{013\}$ or $\{015\}$, both being BIB designs. For most sizes there will be no balanced set and the choice is less clear but might be based on the usual efficiency factor. Combinations of sets provide larger designs and again the question of optimal selection of sets arises. This presents a formidable task for all but small designs. Our principal aim is to show that this task can be greatly simplified if certain isomorphisms between cyclic sets are recognized. A systematic approach for the construction of optimal cyclic designs is then developed.

3. <u>EQUIVALENCE CLASSES</u>. Let us now apply to $\{012\}$ of equation (1) the re-numbering or permutation

$$R(7,\ 3) = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 3 & 6 & 2 & 5 & 1 & 4 \end{pmatrix}$$

obtained by multiplying each of the 7 labels by 3 (mod 7). Then $\{012\}$ becomes

036 362 625 251 514 140 403,

a Youden Type design which is merely a re-arrangement of $\{014\}$. We write $\{012\} \xrightarrow{3} \{014\}$. Thus $\{012\}$ and $\{014\}$ are isomorphic. Two further applications of $R(7, 3)$ give $\{024\}$ and the original $\{012\}$. We have therefore established the equivalence class $\{012\} \sim \{014\} \sim \{024\}$. No blocks need be written in the process if partition notation is used: $\{012\} \xrightarrow{3} \{036\} = (331) = (133) = \{014\} \xrightarrow{3} \{035\} = (322) = (223) = \{024\}$. Likewise $\{013\} \xrightarrow{3} \{032\} = \{023\} = (214) = (142) = \{015\}$, so that $\{013\}$, $\{015\}$ form a second equivalence class.

The same procedure can be used for any prime n and any k. To see this note that the permutations $R(n, 1)$ (the identity permutation), $R(n, 2), \ldots, R(n, n-1)$, form a group under "multiplication" $*$ defined by

$$(2) \qquad R(n,\ i) * R(n,\ j) = R(n,\ ij \bmod n)$$

which is isomorphic with the multiplicative group of residues mod n. Hence all elements $R(n, i)$ are generated by powers of $R(n, g)$, where g is a primitive root of n (i.e., $g^x \not\equiv 1 \bmod n$ for $x = 1, 2, \ldots, n-2$ but

$g^{n-1} \equiv 1 \bmod n$). But a permutation σ which changes one cyclic set into another must be of the form R(n, i) if we assume without loss of generality that σ leaves 0 unchanged; for if a, b, c, d, are elements of the residue set with a and b = a+d two elements in the same block we require that

$$\sigma(b) - \sigma(a) = \sigma(d) \qquad \text{all } a, b, d$$

or

$$\sigma(a) + \sigma(d) = \sigma(a+d),$$

showing that σ is multiplicative: $\sigma(a) = ca$. Thus all possible isomorphisms between cyclic sets can be established conveniently by repeated application of R(n, g).

When n is not prime the R(n, i) continue to form a group under * of (2) provided i and j are restricted to be integersrelatively prime to n. The group is now of order $\phi(n)$ and is clearly isomorphic with the multiplicative group of the reduced set of residues. g is said to be a primitive root of n if $\phi(n)$ is the smallest power making $g^{\phi(n)} \equiv 1 \bmod n$. Primitive roots exist only if n equals 2, 4, p^n, or $2p^n$, where p is any prime > 2 and n any integer. For values of n admitting a primitive root we proceed as before; otherwise, multiplication by each member of the reduced set of residues will establish most isomorphisms.

Example 1 (cont'd.) Since 8 does not have a primitive root we begin by applying R(8, 3) to the sets of Example 1 and find

$$(1115) \xrightarrow{3} (1232), \quad (1124) \xrightarrow{3} (1223), \quad (1142) \xrightarrow{3} (1322).$$

The other sets are unchanged by the transformation. Likewise R(8, 5) gives

$$(1115) \xrightarrow{5} (1232), \quad (1124) \xrightarrow{5} (1322), \quad (1142) \xrightarrow{5} (1223).$$

R(8, 7) produces "mirror images" obtained by reading a circular partition anti-clockwise rather than clockwise. E.g. $(1124) \xrightarrow{7} (4211) = (1142)$. This isomorphism had already been established by R(8, 3) and R(8. 5) because $5 \equiv -3$. However, an additional isomorphism can be obtained by the permutation

$$\binom{0}{0} \quad \binom{1}{1} \quad \binom{2\ 6}{6\ 2} \quad \binom{3\ 7}{7\ 3} \quad \binom{4}{4} \quad \binom{5}{5}$$

which takes (1133) into (1214). This is the only instance we have come across where the equivalence of two cyclic sets cannot be demonstrated by a multiplicative permutation.

A listing of all equivalence classes for cyclic sets in experiments with $n \leq 15$ and $k = 3, 4, 5$, is given in [12]. The efficiences of these sets regarded as designs have also been tabulated. When $n = 8$, $k = 4$ we find

Design			E	E_1	E_2	E_3	E_4
{0123}	=	(1115)	.812	.922	.834	.760	.712
{0124}	=	(1124)	.851	.867	.873	.810	.868
{0125}	=	(1133)	.851	.867	.809	.867	.877
{0134}	=	(1214)	.836	.863	.810	.869	.807
{0145}	=	(1313)	.779	.802	.803	.668	.800 ($r = 2$).

Here E is the overall efficiency and E_j ($j = 1, 2, 3, 4$) is the efficiency factor relating to the comparison of j-th associates. On the basis of E the choice of optimal design for $r = 4$ among the five sets (the fifth duplicated) lies between {0124} and {0125}, with the latter preferable in having only 3 associate classes. It should be noted that except for fully balanced designs the highest value of E does not necessarily correspond to the design with the smallest number of associate classes. Other optimality criteria might be used but the choice of cyclic design is in any case reduced to one of the non-isomorphic sets. Moreover, it is only combinations of these sets (and possible disconnected sets) which need to be considered in the construction of larger cyclic designs. In Table 1 we list the most efficient cyclic sets for $n \leq 15$ and $k = 3, 4, 5$.

Cyclic sets with two associate classes. For purposes of comparison we have made a corresponding compilation in Table 2 of two-associate PBIB designs of all types as given by Bose et al. [1] and (with asterisks) by Clatworthy [4]. The BIB designs in this range are also included. It will be noted that Table 2 has gaps for several (n, k) combinations

although the symmetrical case is favorable to the existence of designs with a high degree of balance. The table also shows that a cyclic design with more than two associate classes may be more efficient than any two-associate PBIB.

It is of some interest that every regular (R) group divisible PBIB of Table 2 may be laid out as a cyclic design; this is already done in [1] in some cases and may be effected for the remaining designs by suitable relabeling. We find the following isomorphisms:

$n = 6$:	R1 $\sim$	$\{013\}$	,	R2 $\sim$	$\{0124\}$	;		
$n = 8$:	R5 $\sim$	$\{013\}$	,	R108* $\sim$	$\{01235\}$	,	R109* $\sim$	$\{01246\}$;
$n = 9$:	R8 $\sim$	$\{0136\}$	,	R112* $\sim$	$\{01346\}$	;		
$n = 10$:	R114* $\sim$	$\{01257\}$	;					
$n = 12$:	R15 $\sim$	$\{0137\}$	,	R116* $\sim$	$\{01356\}$	,		
	R117* $\sim$	$\{01249\}$	,	R118* $\sim$	$\{0147\underline{10}\}$	;		
$n = 14$:	R24 $\sim$	$\{0146\}$	;					
$n = 15$:	R27 $\sim$	$\{0137\}$	.					

There are only two other cyclic designs with two associate classes in the range under consideration. For $n = 13$ we have C1 $\sim$ 014 ; for $n = 12$ the design $\{01247\}$ has the same association scheme as R116* but is not isomorphic with it.

4. <u>COMBINATIONS OF CYCLIC SETS</u>. Cyclic sets for given n may be combined to produce a wide variety of cyclic designs, still of PBIB form. This can always be done if the number of replications r is a multiple of k but will also be possible for certain other values of r if fractional sets exist. We shall say that the combined design is of size (n, k, r). Equivalence classes may again be established. However, the most efficient cyclic design of given size is not necessarily one made up of the most efficient cyclic sets.

Example 2. For $n = 9$, $k = 3$ we have the equivalence classes

$$A : (117), (225), (144);$$

$$B : (126), (243), (153), (162), (234), (135);$$

$$C : (333) \qquad (r = 1).$$

The order within a class has been arranged so that successive sets are obtained by the application of $R(9, 2)$, the primitive root of 9 being 2. There are clearly two non-isomorphic designs of size (9, 3, 4) obtained by combining (333) with any member of class A or class B. Of these the latter, which may be written as $\{013, 036\}$, is the more efficient, with $E = 0.713$ and 4 associate classes.

To get designs with $r = 6$ we can take two sets from A, two from B, or one from each. Call the sets A_1, A_2, A_3, and $B_1, B_2, \ldots, B_6$. We then have the following seven equivalence classes;

$$A_1A_2, A_2A_3, A_3A_1;$$

$$B_1B_2, B_2B_3, B_3B_4, B_4B_5, B_5B_6, B_6B_1;$$

$$B_1B_3, B_2B_4, B_3B_5, B_4B_6, B_5B_1, B_6B_2;$$

$$B_1B_4, B_2B_5, B_3B_6;$$

$$A_1B_1, A_2B_2, A_3B_3; A_1B_4, A_2B_5, A_3B_6;$$

$$A_1B_2, A_2B_3, A_3B_4, A_1B_5, A_2B_6, A_3B_1;$$

$$A_1B_3, A_2B_4, A_3B_5, A_1B_6, A_2B_1, A_3B_2.$$

Calculations show that the most efficient cyclic design is A_1A_2 with $E = 0.731$ and 4 associate classes.

The present example has been chosen to bring out the enumeration procedure required when the original cyclic sets fall into several equivalence classes.

Actually, for $r = 6$ as many as four PBIB(2) designs are available, viz. SR13, R10, LS3, and LS9*, of which LS3 is the most efficient having $E = 0.741$. When $r = 4$ the only tabulated PBIB(2) design is LS6, with the relatively low efficiency $E = 0.667$. For $r \leq 10$ Table 3 lists a selection of cyclic designs in cases where no such PBIB(2) designs are known to exist or are all of more than trivially inferior efficiency.

It is of interest to note that the number of non-isomorphic designs made up of s sets all chosen from the same class of S sets is just $N(s, S-s)$, where N is defined by (2). This is so because we can now regard the beads of Fig. 1 as representing sets rather than blocks. The operation $R(n, g)$, where g is a primitive root, produces a unit turn. The enumeration of non-isomorphic designs when sets are from more than one class proceeds exactly as described in [7] for $k = 2$.

5. FRACTIONAL SETS. The number nk of observations required for a cyclic set of size (n, k) will often be greater than desired, especially when n is large. In this situation fractional sets are very useful. As pointed out in Example 1 such sets are characterized by a repetitive pattern in their partition representation. No such design is possible if n is prime. For n composite fractional sets exist corresponding to every divisor d $(1 < d < n)$ of n since there must be at least one partition of n consisting of d repetitions. Clearly, k must be a multiple of d, and $r = k/d$; (however, $r = 1$ gives a disconnected set). From a cyclic set with parameters $(n/d, k/d)$ a fractional set with parameters $(n, k, r = k/d)$ can always be obtained.

Example 3. For $n = 30$ connected fractional sets exist for $k = 4, 6, 8, 9, 10, \ldots$. Suppose we require a design with $k = 6$. The non-isomorphic connected cyclic sets of size (15, 3) are $(11\underline{13})$, $(12\underline{12})$, $(13\underline{11})$, $(14\underline{10})$, and (159). Of these $(12\underline{12})$ leads to the most efficient design of size (30, 6, 3), viz. $(12\underline{12}12\underline{12})$ or $\{0\,1\,3\,\underline{15}\,\underline{16}\,\underline{18}\}$ with $E = 0.762$.

In [12] a selection of the most efficient fractional sets of given size is tabulated for $n \leq 100$.

6. ACKNOWLEDGMENT. Much of this work was done while the authors were at the Virginia Polytechnic Institute. We are grateful to Dr. Dale M. Mesner for several helpful comments.

REFERENCES

[1] Bose, R. C., Clatworthy, W. H., and Shrikhande, S. S. (1954). Tables of partially balanced designs with two associate classes. North Carolina Agric. Expt. Sta., Tech. Bull. No. 107.

[2] Bose, R. C. and Nair, K. R. (1939). Partially balanced incomplete block designs. Sankhyā, 4 337-372.

[3] Bose, R. C. and Shimamoto, T. (1952). Classification and analysis of partially balanced designs with two associate classes. J. Amer. Statist. Assoc. 47 151-190.

[4] Clatworthy, W. H. (1956). Contributions on partially balanced incomplete block designs with two associate classes. Nat. Bur. Stand., Applied Mathematics Series No. 47.

[5] Das, M. N. (1960). Circular designs. J. Indian Soc. Agric. Statist. 12 46-56.

[6] David, H. A. (1963). The structure of cyclic paired-comparison designs. J. Austral. Math. Soc. 3 117-127.

[7] David, H. A. (1965). Enumeration of cyclic paried-comparison designs. Amer. Math. Monthly, 72 (In press).

[8] Jablonski, E. (1892). Théorie des permutations et des arrangements circulaires complets. J. Math. Pure Appl., 4th series, 8 331-349.

[9] Kempthorne, O. (1953). A class of experimental designs using blocks of two plots. Ann. Math. Statist. 24 76-84.

[10] Masuyama, M. (1962). On the classes of blocks. Rep. Stat. Appl. Res., JUSE 9 83-87.

[11] Pearce, S. C. (1963). The use and classification of non-orthogonal designs. J. Roy. Statist. Soc. A 126 353-369.

REFERENCES (cont'd.)

[12] Wolock, F. W. (1964). Cyclic designs. Ph.D. dissertation, Virginia Polytechnic Institute.

[13] Zoellner, J. A. and Kempthorne, O. (1954). Incomplete block designs with blocks of two plots. Iowa Agric. Expt. Sta., Res. Bull. No. 418.

Table 1. Most efficient symmetric cyclic PBIB design D for n treatments and block size k, and its efficiency E.

n	k=3 D	k=3 E	k=4 D	k=4 E	k=5 D	k=5 E
6	$\{013\}^2$	·784	$\{0123\}$	·895	$\{01234\}^1$	·961
7	$\{013\}^1$	·778	$\{0124\}^1$	·876	$\{01234\}$	·932
8	$\{013\}^2$	·748	$\{0125\}$	·851	$\{01235\}^2$	·914
9	$\{013\}$	·722	$\{0134\}$	·836	$\{01235\}$	·898
10	$\{013\}$	·700	$\{0125\}$	·823	$\{01245\}$	·888
11	$\{013\}$	·676	$\{0125\}$	·817	$\{01247\}^1$	·880
12	$\{014\}$	·673	$\{0137\}^2$	·813	$\{01247\}^2$	·870
13	$\{014\}^2$	·667	$\{0139\}^1$	·812	$\{01269\}$	·863
14	$\{014\}$	·670	$\{0146\}^2$	·805	$\{01358\}$	·859
15	$\{015\}$	·641	$\{0137\}^2$	·795	$\{0124\underline{10}\}$	·853

N. B. Superscripts $^{1,\,2}$ denote respectively BIB and PBIB(2) designs.

Table 2. Balanced (BIB) and two-associate PBIB designs with codes and efficiencies from Bose et al. [1] and Clatworthy* [4].

n	k=3 D	k=3 E	k=4 D	k=4 E	k=5 D	k=5 E
6	R1	.78	S2, R2	.88, .89	BIB	.96
7	BIB	.78	BIB	.88		
8	R5	.75	SR7	.84	R108*, R109*	.91, .90
9	SR12	.73	R8, LS1	.80, .83	LS10, R112*	.90, .89
10	T6	.70	S17, T2	.79, .79	R114*	.88
11			T12	.82	BIB	.88
12			R15	.81	R116*, R117* R118*	.87, .87 .81
13	C1	.67	BIB	.81		
14			R24	.80		
15	T28	.66	R27	.80		

Table 3. Selected cyclic designs with $r > k$, corresponding optimal two-associate PBIB designs, and efficiencies E.

Size (n, k, r)	Cyclic design	E	PBIB(2) design	
8, 3, 6	{013, 014}	·756	R50*	·747
8, 4, 5	{0134, 0246}	·850	-	
9, 3, 4,	{013, 036}	·713	LS6	·667
10, 4, 6	{0147, 0156}	·825	T3	·789
10, 4, 8	{0126, 0148}	·830	R14	·823
11, 3, 6	{013, 026}	·727	-	
11, 3, 9	{013, 014, 027}	·730	-	
11, 4, 8	{0134, 0248}	·823	-	
13, 4, 8	{0125, 0159}	·807	C2	·797
13, 5, 10	{01247, 01258}	·865	-	
14, 3, 9	{014, 02$\underline{11}$, 019}	·709	-	
14, 5, 10	{0124$\underline{10}$, 017$\underline{10}$ $\underline{12}$}	·862	-	
15, 3, 4	{015, 05$\underline{10}$}	·682	T23	·673
15, 5, 6	{01257, 0369$\underline{12}$}	·856	T38*	·808

SOME RESULTS ON THE FOUNDATIONS OF STATISTICAL DECISION THEORY

Bernard Harris, J. D. Church, and F. V. Atkinson
Mathematics Research Center, U. S. Army
The University of Wisconsin

INTRODUCTION. A fundamental problem in statistical decision theory is concerned with establishing criteria for selecting a single decision procedure from the set of available decision procedures. In this paper, some criteria for optimality of statistical decision procedures are proposed and the consequences of these criteria are discussed. It is shown that these optimality criteria exclude a very general class of decision criteria, which contain as members, the minimax and minimax regret criteria. Finally, we note that these optimality conditions are consistent, in that there exists a decision procedure which satisfies all conditions, and a constructive procedure is given for determining such a decision procedure.

THE GENERAL STATISTICAL DECISION PROBLEM. A statistical decision problem is characterized by a set of states of nature S, whose elements will be denoted by s, and a set of pure (non-randomized) decisions D, whose elements will be denoted by d. The statistician selects an element d from D, and if nature is in state s, a loss $L(d, s)$ is incurred. An experiment is conducted and random variables $X_1, X_2, \ldots, X_N$ are observed where $X_1, X_2, \ldots, X_N$ has the probability distribution $P(x_1, x_2, \ldots, x_N \mid s)$. We require that the distributions $P(x_1, x_2, \ldots, x_N \mid s)$ be distinct for every $s \in S$. Then, since the decision is to be made following the experiment, the decision procedure is a function δ from the sample space to the space of decisions D. Let Δ be the set of such functions and note that d is then a random variable, i.e. $d = \delta(X_1, X_2, \ldots, X_N)$. This risk function $\rho(\delta, s)$ is then defined by

$$E[L(d, s)] = \rho(\delta, s).$$

The statistician's objective is to choose δ, so that $\rho(\delta, s)$ is small in some appropriate sense. It will frequently be desirable (in the sense of reducing risk) to augment the set of decisions to include the randomized decisions; and equivalently to augment the set of decision procedures Δ to Φ the set

of randomized decision procedures, whose elements will be denoted by ϕ. Φ is the set of all probability mixtures of elements of Δ.

The fundamental problem of statistical decision theory is to decide how to choose an element $\phi \epsilon \Phi$. We can interpret this as consisting of two sub-problems.

1. What conditions should be imposed on a randomized strategy $\phi \epsilon \Phi$, so that we can regard strategies having those properties as being optimal?

2. Having decided which conditions are appropriate, how do we determine which elements $\phi \epsilon \Phi$ satisfy those conditions? Note that for some sets of possible conditions which one may wish to consider, it may happen that there are no strategies in Φ which satisfy them.

We will make the formal assumption that, in advance of the experiment, the statistician is in "complete ignorance" of which element s of S has been selected by nature. That is, that there is no a priori information available concerning the mechanism by which nature will select an element $s \epsilon S$.

The results stated in the succeeding sections have been established under the following hypotheses.

1. S and D are finite sets, i.e. $S = (s_1, s_2, \ldots, s_n)$, $D = (d_1, d_2, \ldots, d_r)$

2. With probability one, the random vector $(X_1, X_2, \ldots, X_N)$ assumes only a finite number of values.

As a consequence of the two hypotheses stated above, Δ is a finite set, and we can label its elements as $\delta_1, \delta_2, \ldots, \delta_m$.

Despite the restrictive nature of these assumptions, there are a substantial number of statistical problems to which they are applicable, and in addition, many problems may be approximate by problems satisfying the above hypotheses. As an example of a problem which satisfies the above restrictions, consider the following illustration.

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed random variables with

$$P\{X_i = 1\} = p_j, \quad P\{X_i = 0\} = 1 - p_j, \quad 0 < p_j < 1$$

for $i = 1, 2, \ldots, N$; $j = 1, 2$; and $S = \{1, 2\}$. Then, the sample space has 2^n elements. If we let $D = \{1, 2\}$, then Δ consists of all functions from the sample space to D, and hence Δ has 2^{2^n} elements. Hence, for this problem the above assumptions are all satisfied.

We can make this illustration more concrete by noting that the above is essentially the problem of testing whether a coin is fair ($p_1 = \frac{1}{2}$) or has probability ($p_2 = \frac{3}{4}$) of landing heads. We can interpret the two elements of D as being 1: Accept the hypothesis that $s = 1$, i.e. $p = \frac{1}{2}$; 2: Accept the hypothesis that $s = 2$, i.e. $p = \frac{3}{4}$. Thus, the illustration given is an "abstraction" of a test of a simple hypothesis against a simple alternative in a coin tossing problem.

It is well-known, that as a consequence of the above two assumptions we can identify the selection of a decision procedure ϕ with the selection of a point in a convex polyhedron C in Euclidean n-space, where C is generated as the convex hull of the points $(\rho(\delta_1 s_1), \rho(\delta_1 s_2), \ldots, \rho(\delta_1 s_n))$, $\delta \epsilon \Delta$. If we define the matrix A, whose elements are a_{ij}, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$ by $\rho(\delta_i, s_j) = a_{ij}$, then $C = C(A)$, the convex hull of the row vectors of A. Thus, we can use the natural relationship between the matrix A and the polyhedron $C(A)$, and freely characterise all relevant aspects of the problem in terms of either the matrix or the associated polyhedron. The reader is referred to the book by D. Blackwell and M. A. Girshick [1] for the relevant details.

We now turn to the characterization of desirable properties for decision procedures.

THE CHOICE OF A DECISION PROCEDURE. It is convenient at this time to introduce some definitions which will be needed in order to specify those properties of a decision procedure which will be considered desirable.

Definition 1. Two decision procedures ϕ_1, ϕ_2, in Φ will be said to be equivalent if

$$\rho(\phi_1, s_j) = \rho(\phi_2, s_j) \text{ for } j = 1, 2, \ldots, n$$

Definition 2. ϕ_1 is said to be dominated by ϕ_2 if

$$\rho(\phi_2, s_j) \leq \rho(\phi_1, s_j), \quad j = 1, 2, \ldots, n$$

with strict inequality holding for at least one j.

Note that if ϕ_1 is dominated by ϕ_2, then regardless of which state of nature s_j has been selected by nature, the risk using ϕ_1 is always at least as large as that using ϕ_2, and hence ϕ_2 is always to be preferred over ϕ_1.

Definition 3. A decision procedure ϕ_0 is said to be admissible if it is not dominated by any element $\phi \in \Phi$.

Since we have previously noted that dominated strategies are not desirable, then clearly the selection of a strategy should be made from among those that are admissible.

Definition 4. A decision procedure ϕ_0 is essential if it is admissible and if for every pair of decision procedures ϕ_1, $\phi_2 \in \Phi$, with ϕ_1 not equivalent to ϕ_0, and for every real number λ, $0 < \lambda < 1$,

$$\rho(\phi_0, s_j) \neq \lambda\rho(\phi_1, s_j) + (1-\lambda)\rho(\phi_2, s_j)$$

for at least one index j, $1 \leq j \leq n$.

The essential decision procedures are those which are admissible and in addition are also extreme points of the convex polyhedron $C(A)$. These decisions can then be used to generate all strategies which one may wish to consider.

The characterization of optimal decision procedures is equivalent to partitioning Φ into two sets, K - the set of decision procedures which are considered optimal, and Φ - K, those which are non-optimal. Equivalenty, we can characterize $Q(A) \subset C(A)$, the set of optimal vectors in Euclidean n-space.

We now propose eight properties which we believe will characterize a satisfactory decision procedure.

1. For every matrix A, Q(A) is a non-empty subset of C(A).

Clearly this condition is essential, since if Q(A) is empty, we have no decision procedures available for use.

2. If A' can be obtained from A by a permutation of the rows and columns of A, then Q(A') can be obtained from Q(A) by applying the permutation on the columns of A to the coordinates of vectors in Q(A).

Condition 2 says that the relabeling of the states of nature, and the (pure) decision procedures available to the statistician should not affect the decision procedure employed.

3. Every decision procedure with a risk vector in Q(A) is admissible.

This condition is just the observation that the only decision procedures that should be considered are the admissible decision procedures.

4. Q(A) is convex.

The motivation for this property is the following. If ϕ and ϕ' are both optimal, i.e. have their risk vectors in Q(A), then every probability mixture of ϕ and ϕ' will also be optimal.

5. If

$$A_1 = \lambda A_0 + \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_n \\ \cdots & \cdots & \cdots & \cdots \\ c_1 & c_2 & \cdots & c_n \end{bmatrix}$$

where λ is a positive real number and the vector $(c_1, c_2, \ldots c_n)$ is an arbitrary real vector, then

$$Q(A_1) = \{\lambda\tilde{x} + \tilde{c}\ ,\ \tilde{x} \in Q(A_0)\ ,\ \tilde{c} = (c_1, c_2, \ldots, c_n)\}\ .$$

This requirement, includes, for example, invariance under the change of units of the loss function. In particular, if $\lambda = 1$ and $C_j = -\min_{1 \le i \le m} \rho(\delta_i, s_j)$, the matrix A_0 is reduced to its regret matrix.

6. If $C(A_1^T) = C(A_2^T)$, where A^T is the transpose of A, and in addition A_1 can be obtained from A_2 by deleting j columns from A_2, then $Q(A_1)$ can be obtained by deleting the corresponding coordinates from every vector in $Q(A_2)$.

Property 6 includes the column duplication property required by other writers, such as J. Milnor [4]. The point of this property, is that under complete ignorance, the decision problem for the statistician is essentially the same in both cases.

7. Let E_A be the submatrix of A corresponding to essential decision procedures in A. Then, if A_1 and A_2 are two matrices with $C(E_{A_1}) = C(E_{A_2})$, we require that $Q(A_1) = Q(A_2)$.

This says that the set of optimal decision procedures should depend only on those pure strategies which are candidates for good strategies. We might note that a risk vector $\tilde{x} \in C(A)$ is an essential strategy if and only if it uniquely minimizes the risk for some a priori distribution on the states of nature.

8. If $\{A_j\}$ is a sequence of matrices with $\lim_{j \to \infty} A_j = A_0$, and $\tilde{x}_j \in Q(A_j)$ for every $j \ge 1$, then every limit point of $\{x_j\}$ is an element of $Q(A_0)$.

This last condition is a continuity requirement. The reader's intuition may be aided by noting, that if one statistical decision problem may be approximated by another statistical decision problem, then this property requires that optimal decision procedures for the first problem are also approximated by the optimal decision procedures for the second problem.

R. D. Luce and H. Raiffa [3] give an extensive discussion of similar systems of optimal properties. The reader's attention is also specifically directed to papers by H. Chernoff [2] and J. Milnor [4], which deal with this problem.

CONSEQUENCES OF THIS CHOICE OF DESIRABLE PROPERTIES. Let $v_j = \min_{1 \le i \le m} \rho(\delta_i, s_j)$ and define $\tilde{v} = (v_1, v_2, \ldots, v_n)$. Define $||\tilde{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$, $1 \le p \le \infty$, where $||\tilde{x}||_\infty$ is interpreted as $\sup_{1 \le i \le n} |x_i|$. Then, let the class of decision procedures $\Delta_p (1 \le p \le \infty)$ specify as optimal all $\tilde{x} \in C(A)$ which are admissible and satisfy

$$||\tilde{v} - \tilde{x}||_p \le ||\tilde{v} - \tilde{y}||_p$$

for all $y \in C(A)$. Then, the following theorem can be established.

THEOREM. For $1 \le p < \infty$, Δ_p satisfies every property with the exception of property 6. Δ_∞ satisfies every property except property 8.

The reader should note that Δ_1 is Laplace's criterion and that Δ_∞ is the minimax-regret criterion restricted to admissible decision procedures. The failure of the minimax regret criterion to satisfy the above list of properties also establishes that the minimax criterion does not always satisfy the list of requirements given above.

Finally, we have the following theorem.

THEOREM. There is at least one decision procedure satisfying all of the above properties.

The proof of this last statement is accomplished by exhibiting a constructive process, which we now sketch.

Let $\{\epsilon_j\}$, $j = 1, 2, \ldots$ be a monotone non-increasing sequence of positive real numbers, with $\lim_{j\to\infty} \epsilon_j = 0$. Let $d(\widetilde{x}, \widetilde{y}) = \sup_{1 \le i \le n} |x_i - y_i|$ and let $Q_1 = C(A)$. Define $v_i^{(1)} = \min_{\widetilde{x} \in Q_1} x_i$ and $\widetilde{v} = (v_1^{(1)}, v_2^{(1)}, \ldots, v_n^{(1)})$. Then let $z_1 = \min_{x \in Q_1} d(\widetilde{v}_1, \widetilde{x})$. We now proceed inductively. For $h \geq 1$, define

$$Q_{h+1} = \{\widetilde{x} \in Q_h : d(\widetilde{v}_h, \widetilde{x}) \le z_h + \epsilon_h z_1\} \quad \text{where} \quad v_i^{(h)} = \min_{\widetilde{x} \in Q_h} x_i \text{ and}$$

$\widetilde{v}_h = (v_1^{(h)}, v_2^{(h)}, \ldots, v_n^{(h)})$ and $z_h = \min_{\widetilde{x} \in Q_h} d(\widetilde{v}_h, \widetilde{x})$. Then, it can be shown that $Q(A) = \bigcap_{h=1}^{\infty} Q_h$ satisfies all of the requirements.

One of the consequences of the above construction is that $Q(A)$ is a single point $\widetilde{s}$. However, the specific single point obtained may depend on the choice of the sequence $\{\epsilon_j\}$ employed.

The reader's intuition concerning the above construction may be aided by considering the process as a limit of a sequence of minimax-regret procedures, as follows:

z_1 is the minimax regret decision procedure (more properly, it is the distance of the risk vector associated with the minimax regret decision procedure from $\widetilde{v} = \widetilde{v}_1$). Then a new convex polyhedron Q_2, is constructed, containing the risk vector for the minimax regret decision procedure, and the minimax regret decision procedure for Q_2 is determined. The process is repeated and converges to a single point $s^2 = Q(A)$.

REFERENCES

[1] Blackwell, D. and Girshick, M. A. (1954). The Theory of Games and Statistical Decisions. Wiley, New York.

[2] Chernoff, H. (1954). Rational Selection of Decision Functions. Econometrica, 22, 422-443.

[3] Luce, R. D. and Raiffa, H. (1957). Games and Decisions. Wiley, New York.

[4] Milnor, J. W. (1954). Games Against Nature. In Thrall, Coombs, and Davis, Decision Processes, Wiley, New York, 49-60.

PATHOPHYSIOLOGY OF INDIAN COBRA VENOM

James A. Vick, Henry P. Ciuchta, and James H. Manthei
Directorate of Medical Research
US Army Edgewood Arsenal
Chemical Research and Development Laboratories
Edgewood Arsenal, Maryland

INTRODUCTION. It has been reported that the venom of the hooded cobra, *Naja naja*, has a detrimental effect on the respiratory system of animals and man [1-3]. Several workers have attempted to fractionate the crude venom into its various toxic fractions [4, 5], they being: (a) neurotoxic fraction, (b) a cardiotoxic fraction and (c) a non-specific hemolytic fraction. Our study is concerned with the effect of crude cobra venom on cortical electrical activity (EEG), respiration and the cardiovascular system.

MATERIALS AND METHODS. In this study a total of 54 dogs and 5 monkeys of the Cynapthecoid group (sooty mangabey) were used. Of the above total, 44 adult mongrel dogs, anesthetized with sodium pentobarbital, 30 mg/kg, were used to study the effect of venom on the respiratory and cardiovascular systems. Femoral arterial pressure was monitored using a Statham strain gauge and a Grass polygraph recorder. The phrenic nerve was isolated at the level of the 5th cervical vertebra. The nerve was carefully dissected free of connective tissue and sectioned. Silver wire electrodes were connected to the central end of the phrenic from which nerve impulses were then monitored and amplified on a Tektronix oscilloscope. Permanent recordings were obtained photographically. Both EKG and respiratory rate were recorded in some of the animals using a Grass polygraph recorder. All of the above animals were administered (0.5 mg/kg) lyophilized crude cobra venom, which was reconstituted with normal saline and injected directly into the femoral vein.

The 44 animals were divided into the following groups: Group I was comprised of six animals used to study the overall effect of the venom on blood pressure and respiration. Group II - Eight animals ventilated with a Starling pump at respiratory arrest but prior to cardiovascular failure. The resultant effect on survival time was noted. Group III - The remaining thirty animals were used to study specific effects of the

of the venom on the respiratory system, including the phrenic nerve and diaphragm. Nerve impulses over the central end of the cut phrenic nerve were continuously observed. The peripheral end of the cut phrenic nerve and the diaphragm were stimulated at intervals using a Grass model 54 stimulator. Diaphragmatic muscle contractions were recorded with a Grass Force Displacement Transducer. The effect of venom, artificial respiration and changes in pO_2 and pCO_2 tension on nerve activity were observed. Group IV comprised the remaining ten dogs and five monkeys which were used to monitor the effect of crude cobra venom (0.5 mg/kg) on cortical electrical activity. Blood pressure and respiratory effects were also recorded. The cortical electrical activity was recorded using bipolar silver electrodes which were surgically implanted directly on the dura of each hemisphere of the brain. Continuous electroencephalograms were recorded prior to and for up to 10 hours after the intravenous administration of the venom.

RESULTS.

Group I. The effect of venom on respiratory rate and arterial blood pressure is shown in Figure 1. Within 1-5 minutes post-injection there is an increase in respiratory rate as well as a sharp drop in blood pressure. This is followed by a progressive decrease in respiratory rate and volume to complete arrest at 90-120 minutes. During this time blood pressure makes a partial recovery remaining stable until respiratory failure, at which time cardiovascular collapse results. The average survival time of this group was 105 minutes.

Group II. The effect of venom on the artifically ventilated animal is shown in Figure 2. These animals were placed on a positive pressure respirator at time of respiratory cessation, with a resultant increase in survival time of from 4-6 hours. However, all animals untimately developed arrhythmias and progressive hypotension which led to death Figure 3. The average survival time for this group of animals was 7.5 hours post-venom.

Group III. Changes in phrenic nerve action potentials induced by cobra venom are shown in Figure 4. Action potentials prior to venom are synchronous corresponding to the inspiratory phase of respiration. Increase in both rate and amplitude are noted within 1-5 minutes after

administration of venom. The central component of the nerve continues to discharge for from 5-10 minutes after complete cessation of respiration. During this period phasic discharges over the phrenic nerve become sporadic and irregular. These central impulses are eliminated by placing the animal on the artificial respirator. At any time prior to death impulse traffic can again be re-established by discontinuing artificial respiration, even though the animals do not breathe spontaneously. Phrenic impulses, as seen on the oscilloscope, continue with increasing frequency and amplitude until the animal either expires or is again ventilated.

The administration of 5 percent CO_2 to artifically ventilated animals initiates discharges over the phrenic nerve. This is quickly eliminated by removal of the stimulus. Phasic phrenic discharges can also be elicited in ventilated animals by the reduction of their tidal volume. Where such activity is noted the administration of 100 percent oxygen does not eliminate or appreciably alter their frequency or amplitude.

The terminal effect of venom on impulse traffic over the phrenic nerve is characterized by abnormal appearing bursts probably due to a combination of hypotension and central nervous system ischemia.

Spontaneous contractions of the diaphragm show a gradual decrease in force of contraction after venom ultimately leading to complete cessation of movement Figure 5 [6].

Group IV. The effect of crude cobra venom (0.5 mg/kg) on the EEG of the dog and monkey can be seen in Figure 6. Within 30-60 seconds following the administration of the venom there was complete loss of EEG, as well as corneal reflexes. There also occurred a sharp drop in arterial blood pressure shortly after cessation of all EEG activity. This hypotension was followed by a partial recovery. The effect of the venom on EEG was irreversible. As seen in Table I all animals expired, with an average survival time of 1.4 hours in the dog and 2.0 hours in the monkey.

DISCUSSION. This study has characterized the effects of crude cobra venom (0.5 mg/kg) on the peripheral respiratory mechanism, cardiovascular system and cortical electrical activity (EEG) of the dog and monkey. The respiratory effect is apparently due to a blockage of nerve impulses at the neuromuscular junction of the diaphragm. This

is supported by the fact that the respiratory center remains functional after venom. There are continued phrenic dischagres, although somewhat modified following the venom. The muscle of the diaphragm remains in tact in that it retains its response to stimuli. This same stimulation when applied to the phrenic nerve produces no response in the diaphragm. It appears, therefore, that transmission of impulses is interferred with at the level of the neuromuscular junction. The character of this block is unknown.

The primary lethal effect of cobra venom, respiratory arrest, was shown to be alleviated with the application of artificial ventilation. This, however, was a temporary phenomena in that all animals eventually developed cardiovascular failure. The etiology of this phenomenon has not been studied but may be related to the action of venom on motor end plates [7]. The effect of venom on cardiovascular hemodynamics may also be due in part to its strong hemolytic effect, producing a high serum potassium which may result in cardiac failure [6].

The cortical electrical activity of the brain of the dog and monkey has been shown to be severely depressed by the action of cobra venom. The exact action of venom is not clear but may also, in some way, be related to its blocking effect on neuromuscular transmission [8].

SUMMARY. This study has dealt with the effects of cobra venom, Naja naja, on the respiratory system cardiovascular system and the cortical electrical activity of the dog and monkey. Results have indicated that death is primarily due to respiratory failure, which appears due to peripheral neuromuscular blockade. The character of this block is unknown. The respiratory center, phrenic nerve and diaphragmatic muscle fibers appear to be relatively unaffected by the venom. Survival time was increased several hours with artificial ventilation, however, all eventually developed cardiovascular difficulties terminating in death. This effect may be due to the extended action of venom on the areas of the body.

In addition, venom has been shown to have a severe depressant effect on the cortical electrical activity of the dog and monkey. The exact mechanism by which this effect is produced has not as yet been defined.

REFERENCES

1. Ganguly, S.N. and Malkana, M.T.: Indian F. med. Res. 24: 281, 1936.

2. Sarkar, B.B., Mitra, S.R and Ghosh, B.N.: Indian F. med. Res. 30: 453, 1942.

3. Gautrelet, J., Halpern, N. and Cortiggiani, E.: Arch. int. Physiol. 38: 293, 1934.

4. Master, R.W.P. and Rao, S.S.: F. biol. Chem. 236: 1986, 1961.

5. Sarkar, N.K.: Ann. Biochem. exper. Med. 8: 11, 1948.

6. Vick, J.A. Ciuchta, H.P. and Polley, E.H.: Arch. int. Pharmacodyn. Vol. 153, No. 2: 424-429, 1965.

7. Chang, C.C. and Lee, C.Y.: Arch. int. Pharmacodyn. 144: 214, 1963.

8. Vick, J.A., Ciuchta, H.P. and Polley, E.H.: Nature. Vol. 203, No. 4952: 1387-1388, Sept. 26, 1964.

LEGENDS FOR ILLUSTRATIONS

Figure 1. The effect of cobra venom on arterial blood pressure and respiratory rate.

Figure 2. Modification of venom effect by use of artificial respirator.

Figure 3. The effect of cobra venom on cardiovascular function after respiratory arrest and subsequent artifical ventilation.

Figure 4. Changes in phasic phrenic discharges produced by cobra venom. Effects of artificial respiration and administration of 5 percent CO_2 after cessation of spontaneous respiration are shown.

Figure 5. Effect of cobra venom on blood pressure, phrenic nerve discharges and diaphragmatic contractions. Note: Loss of diaphragmatic response to direct phrenic stimulation (PS). Diaphragmatic responses to direct stimulation (DS) are retained.

Figure 6. The effect of cobra venom on EEG and blood pressure.

TABLE I

Effect of Cobra Venom on Cortical Electrical Activity

	No. of animals	EEG change	Average survival time (h)
Dogs	10	10/10	1.4 (0.3-2.2)
Monkeys	5	5/5	2.0 (1.1-3.1)

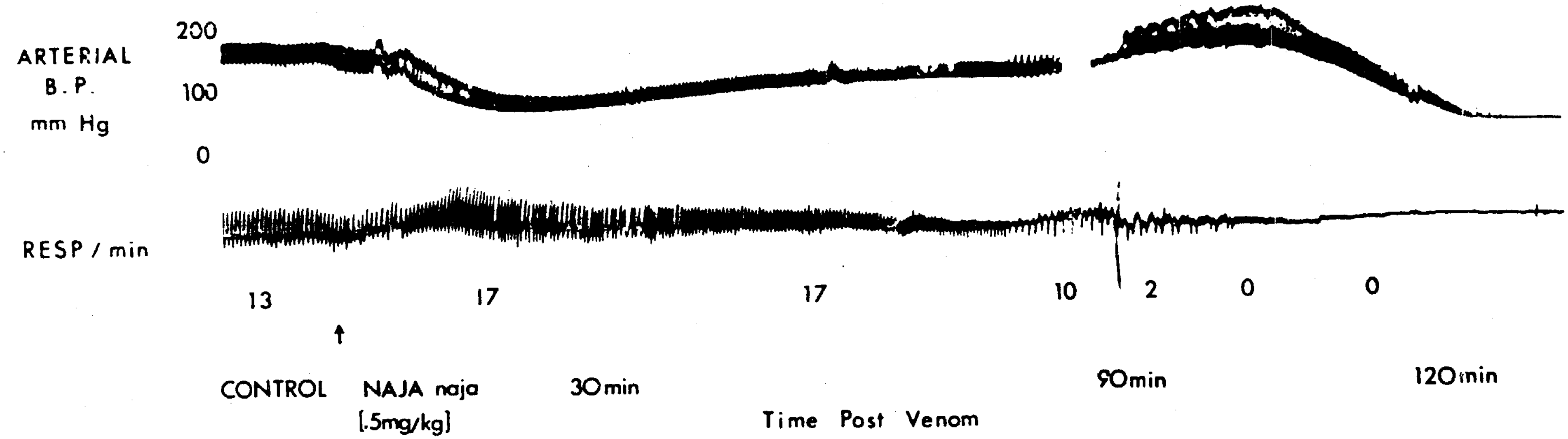
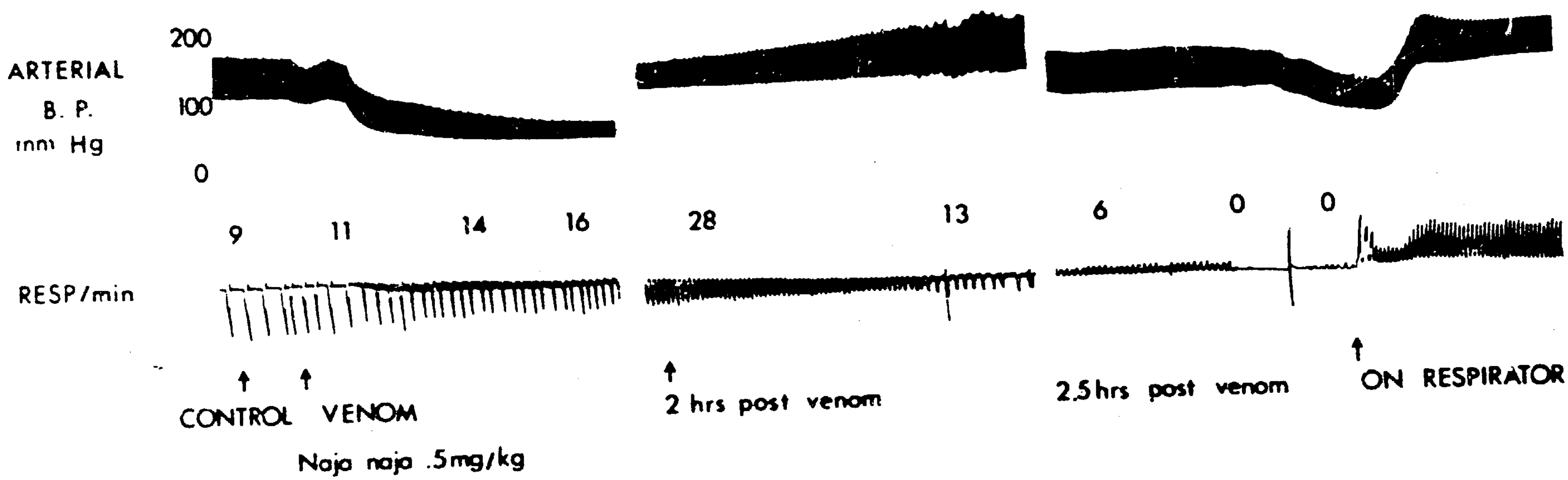
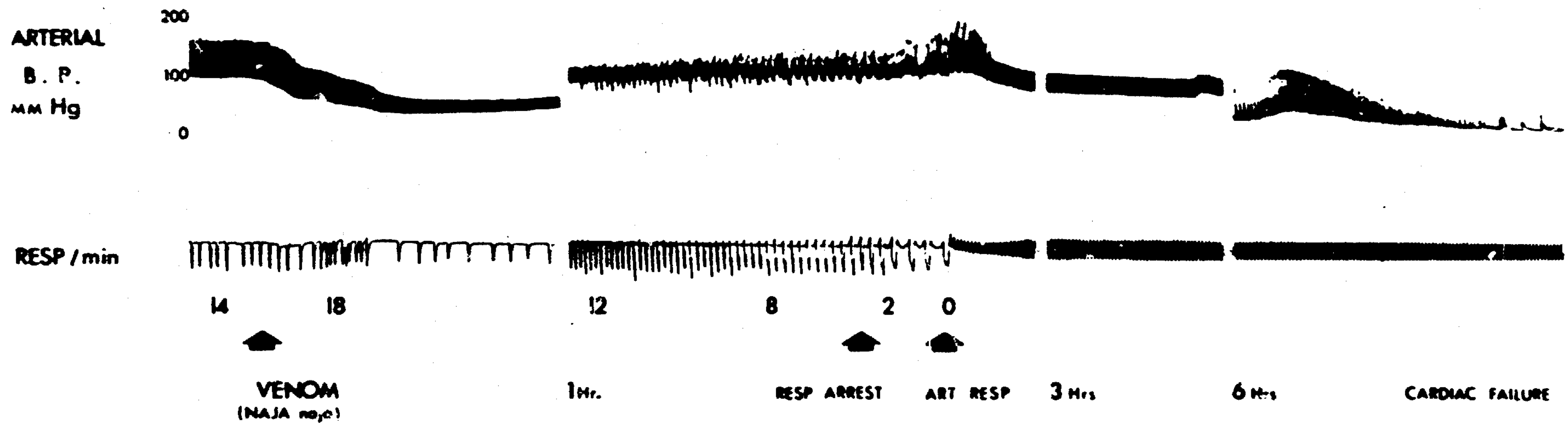

ARTERIAL
B.P.
mm Hg
200
100
0
RESP / min
13
17
17
10
2
0
0
CONTROL
NAJA naja
[.5mg/kg]
30min
90min
120min
Time Post Venom

Figure 1



ARTERIAL
B. P.
mm Hg
200
100
0
9
11
14
16
28
13
6
0
0
RESP/min
CONTROL
VENOM
Naja naja .5mg/kg
2 hrs post venom
2.5hrs post venom
ON RESPIRATOR

Figure 2



ARTERIAL
B. P.
MM Hg
200
100
0
RESP / min
14
18
12
8
2
0
VENOM
(NAJA nojc)
1Hr.
RESP ARREST
ART RESP
3 Hrs
6 Hrs
CARDIAC FAILURE

Figure 3

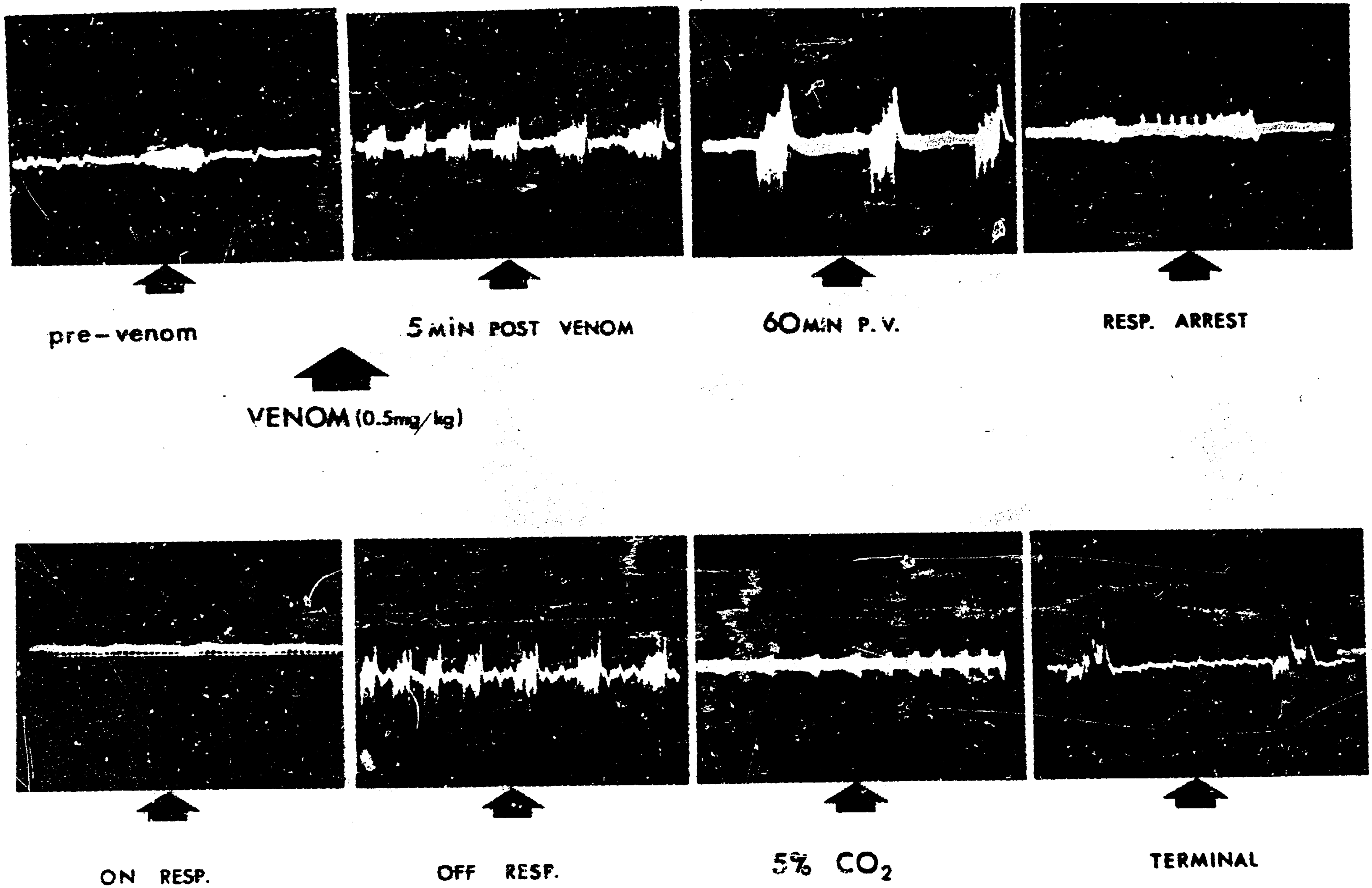

pre-venom
5 MIN POST VENOM
60 MIN P.V.
RESP. ARREST
VENOM (0.5mg/kg)
ON RESP.
OFF RESP.
5% CO_2
TERMINAL

Figure 4

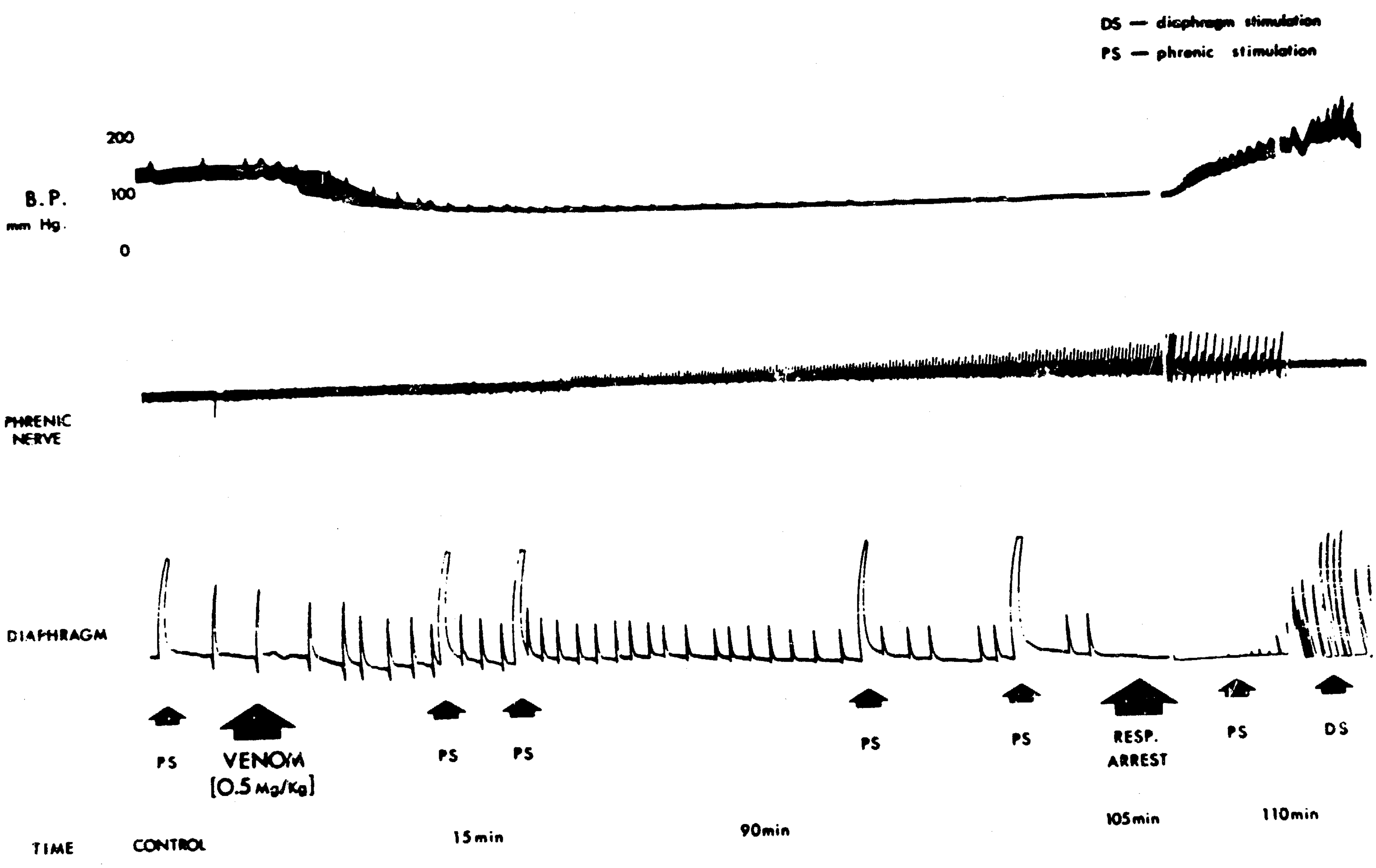

DS — diaphragm stimulation
PS — phrenic stimulation
B.P.
mm Hg.
200
100
0
PHRENIC NERVE
DIAPHRAGM
PS
VENOM
[0.5 Mg/Kg]
PS
PS
PS
PS
RESP. ARREST
PS
DS
TIME
CONTROL
15 min
90min
105min
110min

Figure 5

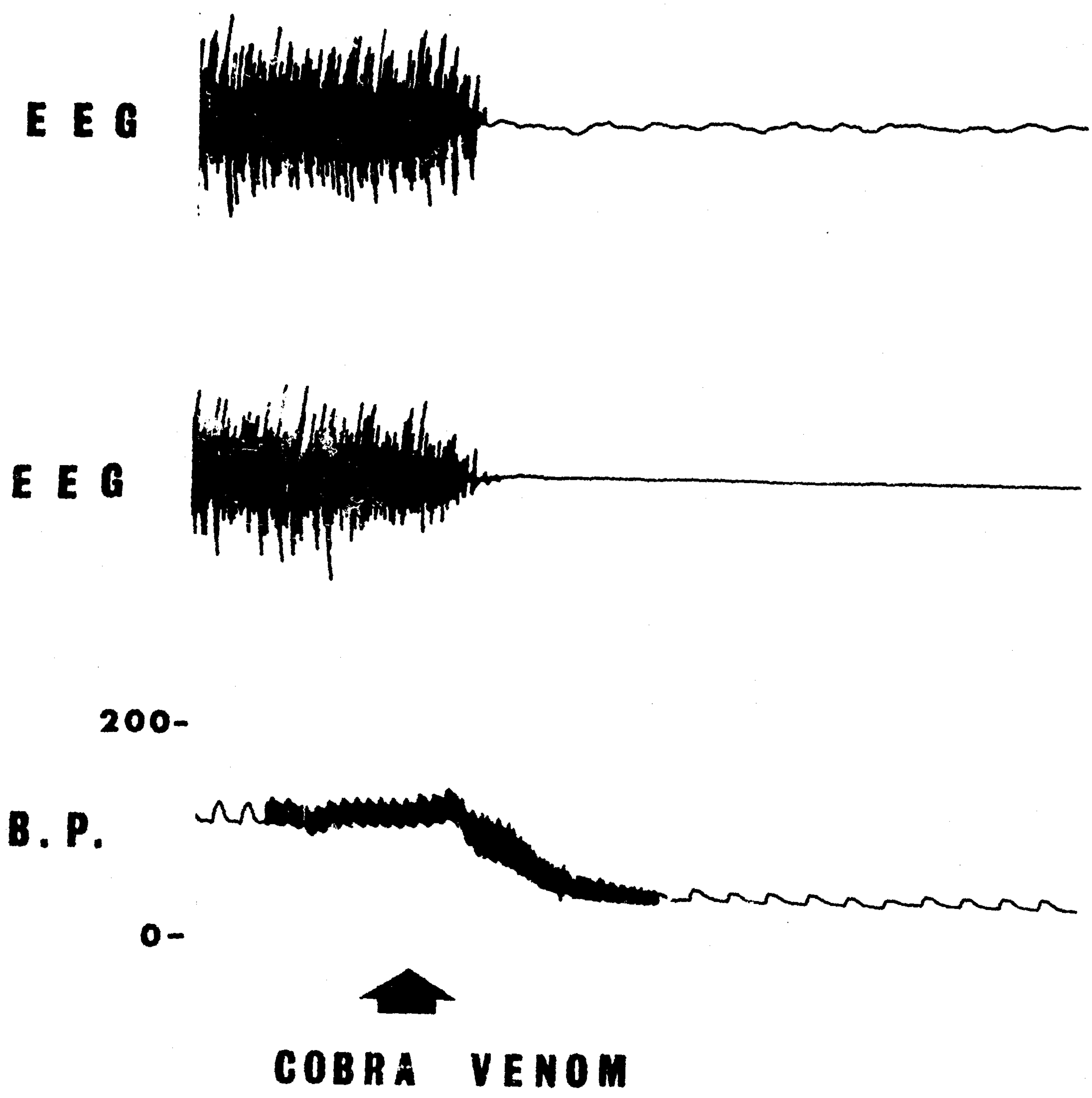


Figure 6

COMPUTER ANALYSIS OF VISUAL DISCRIMINATION DATA

John C. Atkinson
Directorate of Medical Research,
Chemical Research & Development Laboratories
Edgewood Arsenal, Maryland

One of the methods used by the Directorate of Medical Research, Chemical Research and Development Laboratories in evaluating the effect of various drugs on an animal's performance is a visual discrimination test. This is a conditioned visual discrimination between a triangle and a square in which monkeys are trained to avoid or escape an electric shock by pressing a lever under the correct symbol, the triangle. Thus, successful performance involves sensory perception (vision) decision making and motor activity (pressing the lever).

If a drug interfers with any of these activities the result will be a slowed or inaccurate performance. An obvious correlation can be seen between this test and many tasks performed by a soldier during combat.

In our operation Rhesus monkeys are used as test subjects. Each monkey is placed in a sound attenuated booth which is enclosed to prevent visual as well as audio distraction. The monkey is restrained by a Wilinski harness*[1]. By this means the monkey is kept in front of a panel on which there are two screens at an equal level. At the beginning of a trial a triangle appears on one screen and a square on the other. If the monkey presses the lever under the triangle, the symbols disappear from the screen and the trial is over. If he presses the lever under the square he receives a punishment in the form of a mild electrical shock for twenty (20) milliseconds. This is called an incorrect response. If the monkey does not press the designated lever in an interval of five (5) seconds he receives a negative reinforcement in the form of a mild electrical shock. This shock continues for five (5) seconds unless sooner shut off by pressing the lever under the triangle. Pressing the correct lever before the shock is considered an avoidance response. Pressing the correct lever after the shock has started is considered an escape response. Never

---

*Patent Pending

[1] Frank T. Wilinski - Effects of Atropine Sulfate on Trained Monkeys. Manuscript in progress.

pressing the correct lever is a no response. The time interval between the trial start and a correct response is considered response latency. Pressing either lever when their is no figure on the screen is called an intertrial response.

The electrical equipment associated with trial presentation and the paper tape punch are rack mounted behind each booth. Two (2) loops of punched mylar tape on each rack control the presentation of the trials. The shorter loop initiates trial starts and is punched at random intervals in order that no discernible trial start pattern will be presented to the monkey. Circuity in the rack presents the triangle on the right or left screen in a random order with the restriction that the long term expectation of the number of presentations on the two sides be equal. The longer tape starts and stops the trial presentation tape. The monkeys are given five (5) sessions per day, 55 minutes each, with a five (5) minute break between sessions. No presentations are made for the remaining 19 hours. The monkeys live in the test booths for several days during testing. Food and water are provided adlibitum and the cage provides enough room for the monkey to lie down. A tray of absorbent material underneath the woven wire floor is provided for excretions.

A record of the experiment is made on punched paper tape containing six (6) information codes. When a trial is initiated the punch emits a "start of trial punch" and continues to run at 10 characters per second, emitting a code associated with latency until a correct response is made or the trial is automatically terminated. Separate codes are made for latencies following a right or left screen presentation. At present no distinction is made between right or left latencies upon computer analysis. A code is associated with the avoidance response, with the escape response and with an incorrect response. A separate code for right or left presentations is provided for an intertrial response. Since it is quite possible that two or three of the above things could happen in a single 1/10 second period, the code is designed for this. Forty-one separate codes may appear. Since the tape has 6 information channels it is possible to represent up to 64 codes thus 41 presents no coding problem. At the end of a session an end of session code is automatically punched.

At the end of a day the punched paper tapes are removed from the take-up roll on each equipment rack and sent to the computer group for analysis. During the 55 minute sessions an average of 104 trial presentations are made. To obtain frequent measurements of the progress of the monkey,

the data are considered as 4 subgroups by the computer. These subgroups are termed segments. The first 3 segments contain exactly 26 trials while the last contains the number of trials remaining.

For each segment and for the session the geometric mean of the trial latencies is computed. The computer determines each trial latency by counting the number of tape frames between the start of trial punch and an avoidance or escape punch. If neither occur in 100 tape frames this is considered a no response, and the latency is taken as 10 seconds. The standard error is computed in terms of log latencies for each segment and each session. The 95% fiducial limits are computed for the geometric mean latency for each segment and for each session, and the mean and its limits are printed. Analysis in terms of log latencies is done to minimize the skewness of the latencies which results from the physical inability of the animal to react in less than 2 or 3 tenths of a second. This would truncate the deviations on the minus side. Deviations on the positive side are only truncated after the cut off time of 10 seconds. Since the mean response time is generally from 1/2 to 1 second the positive deviations can be many times as large as the negative ones. This causes skewness. Conversion of the latencies to their logarithms minimizes this skewness.

Session one of each day is considered a control run and any drug is administered between session one and two. A "t" test for significance is made between the mean latency in terms of logarithms of each of the 4 subsequent sessions and the control run. The statement "not significant, or significant at 95%, or significant at 99%, or significant at 99.9%" is printed after each of the sessions 2, 3, 4 and 5. The sum of all latencies for a session is printed at the end of each session. In addition to the analysis of the latencies the computer counts and prints for each segment the number of occurrences of each of the following: avoidance responses, escape responses, incorrect responses done with the right hand, incorrect responses done with the left hand, intertrial responses done with the right hand, intertrial responses done with the left hand and the no responses. No analysis is made of these figures at the present time.

A typical computer printed output is presented as Figure I. The numeric portion of this output is simultaneously punched into paper tape. This tape is to be converted into Holorith cards for storage and will allow future manipulation of the test results.

FIGURE 1

COMPUTER OUTPUT FROM VISUAL DISCRIMINATION ANALYSIS

G. MEAN	L. LIMIT	U. LIMIT	A.R.	E.R.	R. INC.	L. INC.	R. INT.	L. INT.	N.R.
.6879	.6033	.7843	26.0000	.0000	.0000	.0000	7.0000	8.0000	.0000
.7576	.6954	.8254	26.0000	.0000	.0000	.0000	5.0000	2.0000	.0000
.7480	.6839	.8181	26.0000	.0000	.0000	.0000	8.0000	.0000	.0000
.7471	.6886	.8106	25.0000	.0000	.0000	.0000	4.0000	2.0000	.0000
.7345	.7001	.7706							80.2998
.8709	.7947	.9544	26.0000	.0000	.0000	.0000	5.0000	4.0000	.0000
.8873	.7980	.9866	26.0000	.0000	.0000	.0000	3.0000	2.0000	.0000
.8404	.7905	.8934	26.0000	.0000	.0000	.0000	4.0000	1.0000	.0000
.9255	.8774	.9763	24.0000	.0000	.0000	.0000	3.0000	2.0000	.0000
.8796	.8455	.9152							93.3999
SIGNIFICANT 99 PER CENT									
.9500	.8715	1.0356	26.0000	.0000	.0000	.0000	4.0000	2.0000	.0000
1.0081	.8948	1.1357	26.0000	.0000	.0000	.0000	12.0000	6.0000	.0000
1.0831	.9571	1.2258	26.0000	.0000	.0000	.0000	7.0000	1.0000	.0000
1.0840	.9932	1.1831	21.0000	.0000	.0000	.0000	9.0000	.0000	.0000
1.0271	.9744	1.0826							107.2998
SIGNIFICANT 99.9 PER CENT									

G. MEAN = GEOMETRIC MEAN
U. LIMIT = UPPER LIMIT (95 PER CENT CONFIDENCE) L. LIMIT = LOWER LIMIT (95 PER CENT CONFIDENCE)
E.R. = ESCAPE RESPONSES A.R. = AVOIDANCE RESPONSES
L. INC. = LEFT HAND INCORRECTS R. INC. = RIGHT HAND INCORRECTS
L. INT. = LEFT HAND INTERTRIALS R. INT. = RIGHT HAND INTERTRIALS
N.R. = NO RESPONSES
NUMBER AT END OF EACH SESSION UNDER NO RESPONSE COLUMN IS TOTAL LATENCY FOR SESSION

FATIGUE-LIMIT ANALYSES AND DESIGN OF FATIGUE EXPERIMENTS

A. H. Soni and R. E. Little
Oklahoma State University
Stillwater, Oklahoma

INTRODUCTION. It is generally accepted that there is as much, if not more scatter associated with fatigue than with any other mode of failure. Consequently, fatigue presents a challenging problem to both the engineer and the statistician.

The purpose of fatigue analyses is to adduce information about the probability of relatively rare events, not to describe the mean or modal event. Accordingly, the statistical problem in fatigue is to establish the alternating stress amplitude that corresponds to the optimum economic level of tolerable failures.

A brief resume of the nature of fatigue is presented here before discussing existing data and the design of future experiments.

NATURE OF METAL FATIGUE. Fatigue is caused by continued cycle stressing. A fatigue failure can be recognized by fitting the two broken pieces back together and observing the original geometry. As indicated in Figure 1, there is no evidence of gross plastic deformation prior to failure by fatigue.

Fatigue cracks are the cumulative result of micro-inelastic behavior occuring within the substructure of the metal. Electron microscopy and X-ray diffraction studies have shown:

(1) the physical mechanisms associated with fatigue are of a 10^{-3} to 10^{-7} cm observation level, and

(2) these physical mechanisms are intimately related to actual defects (dislocations) in the theoretical atomic arrangement.

The statistical nature of fatigue is intuitively apparent when fatigue is viewed as being caused by these minute substructural defects.

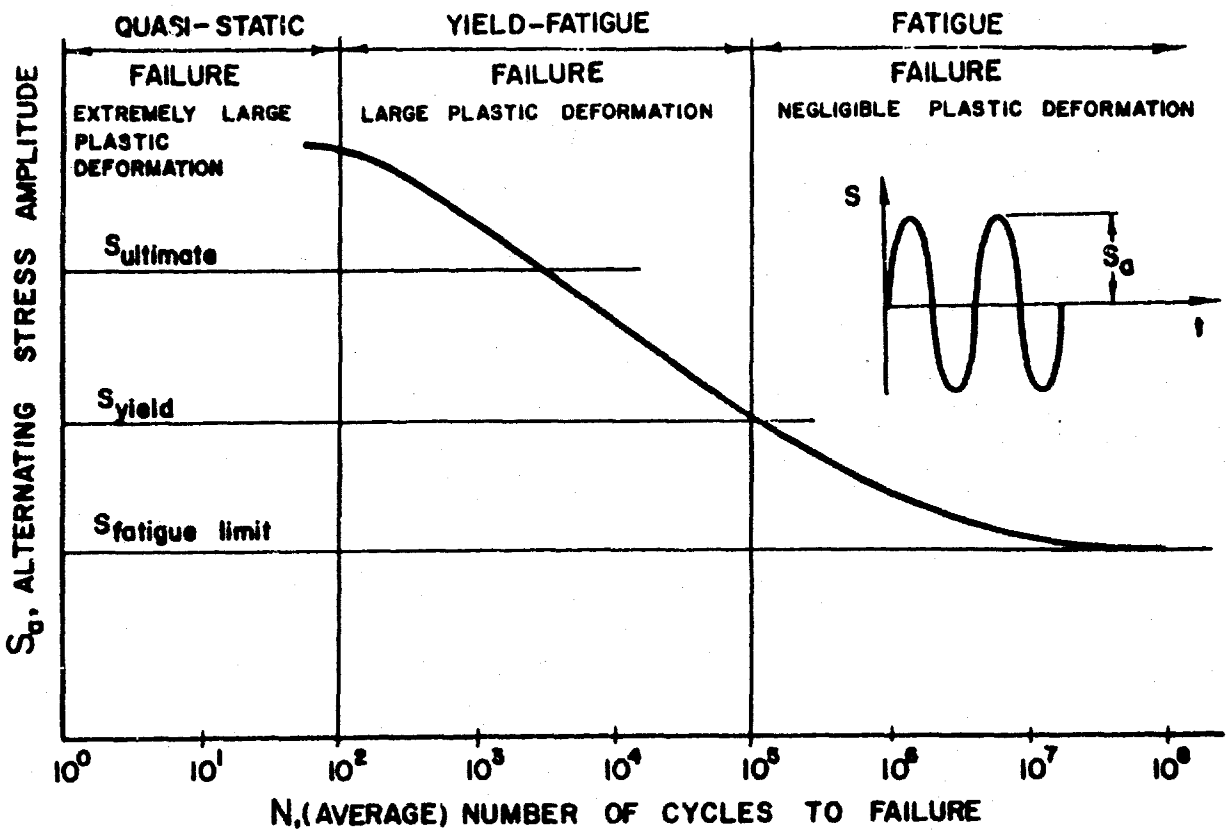


Figure 1 S-N Curve
The lower the alternating stress amplitude the greater the over-all fatigue life, N_f

This intuitive view can be enhanced by considering an idealized material model. First, recall that metals are aggregate structures of randomly oriented crystallites (grains), and that individual crystallites are anisotropic (exhibit differenct properties and strengths in different directions). Now consider the static yield strength of the metallic tensile specimen shown in Figure 2. It theoretically has a unique yield strength only if all crystallites are perfect and have the same orientation. But, since the crystallites of commercial metals have defects and are randomly oriented, the crystallites within this specimen must exhibit a strength distribution.

Observe in Figure 2 that only a few crystallites experience yielding at low stress levels. But, under alternating stressing (Figure 1), these few crystallites yield first in tension, then in compression, then again in tension, and so forth. This localized reversed slip deformation will eventually lead to a fatigue crack in crystallites where the slip magnitude (fatigue intensity) is high. Thus, the number of crystallites that serve as potential fatigue crack initiation sites as well as the fatigue intensity at these sites are directly related to the crystallite strength distribution. Accordingly, fatigue is a statistical problem.

Fatigue failure theories are in their infancy---theory lags experimental work. The present criterion for the relative evaluation of various statistical functions is simply their goodness of fit with regard to data. Figure 3 shows the two types of fatigue data considered, namely:

(1) data stated in terms of a life distribution.

(2) data stated in terms of a strength distribution.

In turn, the over-all objective of all statistical analyses of fatigue data is to develop the P-S-N surface shown in Figure 4.

Existing data indicates that the P-S-N surface is warped and cannot be described in its entirety by a simply mathematical function. This paper treats a small but significant portion of this surface---the statistical analyses of fatigue-limits in terms of a strength distribution.

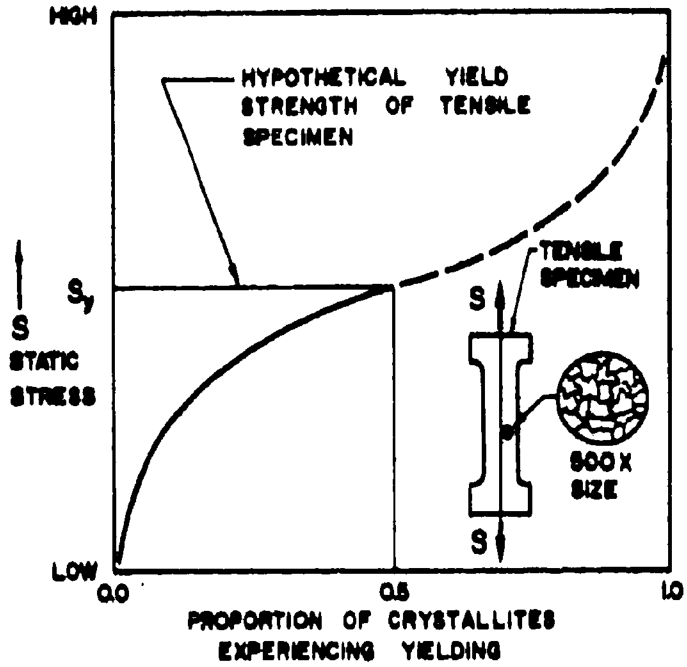


Figure 2 Strength Distribution of Tensile Specimen Crystallites.

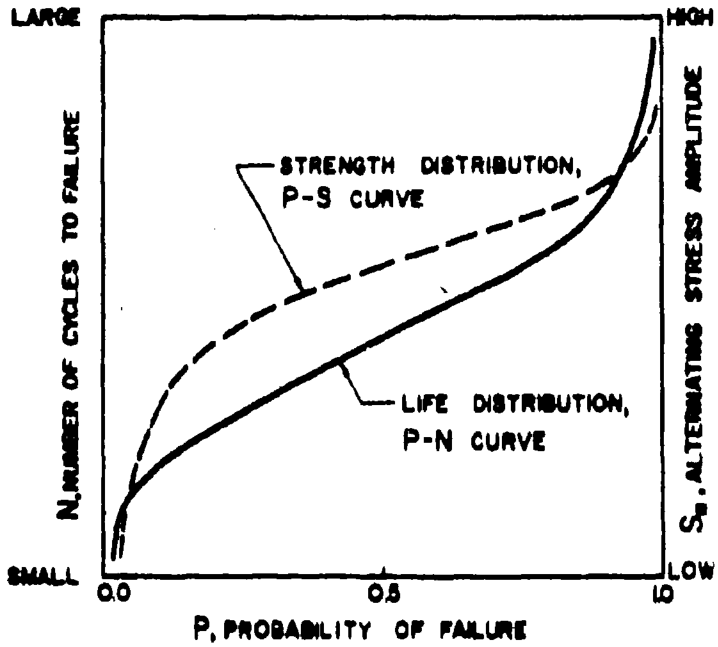


Figure 3 Fatigue Strength and Fatigue Life Distributions
The life distribution is markedly skewed to the right.

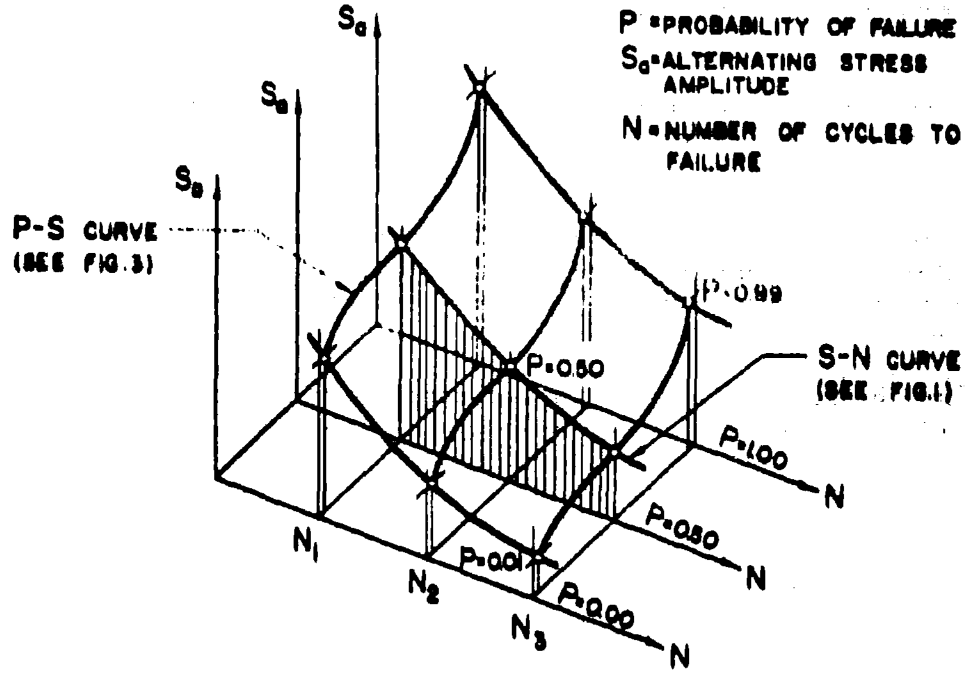


Figure 4 P-S-N Surface
This surface completely defines fatigue failure.

PART I - ANALYSES OF EXISTING FATIGUE-LIMIT DATA

COMMON DISTRIBUTIONS. The three common statistical functions applied herein to fatigue-limit data are listed rows 1, 2, and 3 of Table 1.

Typical fatigue-limit data appears in Table 2. Observe that the statistics recorded are simply the alternating stress amplitudes and the corresponding proportion of specimens failed prior to the given fatigue life.

These common functions are fitted to the observed statistics by using a minimum residual χ^2 approach. For example, the logistic function is fitted by minimizing the logit $\chi^2 = \Sigma\ Npq(\ell - \hat{\ell})^2$, where $\hat{\ell} = \hat{\alpha} + \hat{\beta}\, s$. Taking the partial derivative of the logit χ^2 with respect to $\hat{\alpha}$ and $\hat{\beta}$ and then setting these expressions equal zero; simultaneous solution of the two resulting equations yields the expressions for the estimates listed in rows 4 and 5 of Table 1.

OTHER DISTRIBUTIONS. The goodness of fit of the common (two-parameter) functions can be evaluated by examining the goodness of fit for three-parameter functions, i.e., determining whether the third parameter is really required to describe the data.

Table 3 lists these three-parameter functions. The estimates listed in rows 4, 5, and 6 are established by taking the partial derivative of χ_3^2 with respect to $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$, respectively; setting these expressions equal to zero, and then solving these three equations simultaneously.

The significance of the third parameter, γ, can be now determined from the magnitude of

$$F \approx \frac{\chi^2 - \chi_3^2}{(\chi_3^2 / K-3)}$$

where F has one and (K-3) degrees of freedom. (At least five datum points are desirable for comparative residual χ^2 analyses.)

TABLE 1. COMMON DISTRIBUTIONS

	Normal	Logistic	Extreme Value - Type I
Probability Function	$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} e^{-\frac{t^2}{2}} dt$	$P = \frac{1}{1 + e^{-(\alpha + \beta s)}}$	$P = 1 - e^{-e^{-(\alpha + \beta s)}}$
Linear Transform	$Y = \alpha + \beta s$	$\ell = \ln \frac{P}{Q} = \alpha + \beta s$	$g = \ln\ln \frac{1}{1 - P} = -(\alpha + \beta s)$
Parameters	α, β	α, β	α, β
Estimate of β	$\hat{\beta} = \frac{\Sigma w \Sigma w Y s - \Sigma w Y \Sigma w s}{\Sigma w \Sigma w s^2 - (\Sigma w s)^2}$ Weights, w, are listed in Ref. [4].	$\hat{\beta} = \frac{\Sigma w \Sigma w \ell s - \Sigma w \ell \Sigma w s}{\Sigma w \Sigma w s^2 - (\Sigma w s)^2}$ Where $w = Npq$	$\hat{\beta} = -\frac{K \Sigma g s - \Sigma g \Sigma s}{K \Sigma s^2 - (\Sigma s)^2}$ Where K = No. of Stress Amplitudes
Estimate of α	$\hat{\alpha} = \frac{\Sigma w Y - \hat{\beta} \Sigma w s}{\Sigma w}$	$\hat{\alpha} = \frac{\Sigma w \ell - \hat{\beta} \Sigma w s}{\Sigma w}$	$\hat{\alpha} = -\frac{\Sigma g - \hat{\beta} \Sigma s}{K}$

TABLE 2. RESULTS OF ROTATING BENDING FATIGUE TESTS ON SAE 4340 STEEL. ($N = 10^7$ cycles)

$S_u = 190$ ksi, $K_t = 2.6$.

(Data by Cummings, Stulen, and Schulte)

Test	Stress Level s, ksi	Number Tested	Number Failed	Proportion Failed p
1	32	110	1	0.0091
2	35	60	3	0.0500
3	38	30	6	0.2000
4	41	20	14	0.7000
5	42	20	16	0.8000

TABLE 3. THREE-PARAMETER DISTRIBUTIONS

	Normal	Logistic	Extreme Value--Type I	Weibull (a)
Probability Function	$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} e^{-\frac{t^2}{2}} dt$	$P = \frac{1}{1 + e^{-(\alpha + \beta s + \gamma s^2)}}$	$P = 1 - e^{-e^{-(\alpha + \beta s + \gamma s^2)}}$	$P = 1 - e^{-\frac{(s - \gamma)^{\beta}}{\alpha'}}$
Modified Transform	$Y = \alpha + \beta s_1 + \gamma s_2$ where $s_2 = s_1^2 = s^2$	$\ell = \alpha + \beta s_1 + \gamma s_2$ where $\ell = \ln \frac{P}{1 - P}$; $s_2 = s_1^2 = s^2$	$g = -(\alpha + \beta s_1 + \gamma s_2)$ where $g = \ln\ln \frac{1}{1 - P}$; $s_2 = s_1^2 = s^2$	$g = \alpha + \beta x$ where $\alpha = \ln \alpha'$ and $x = \ln (s - \gamma)$
Parameters	α, β, γ	α, β, γ	α, β, γ	α, β, γ
Estimate of $\hat{\beta}$ (b)	$\hat{\beta} = \frac{\Sigma w S_1 S_2 \Sigma w Y S_2 - \Sigma w Y S_1 \Sigma w Y S_2^2}{(\Sigma w S_1 S_2)^2 - \Sigma w S_1^2 \Sigma w S_2^2}$ Weights, w, are given in ref. [4].	$\hat{\beta} = \frac{\Sigma w S_1 S_2 \Sigma w L S_2 - \Sigma w L S_1 \Sigma w S_2^2}{(\Sigma w S_1 S_2)^2 - \Sigma w S_1^2 \Sigma w S_2^2}$ where $w = Npq$	$\hat{\beta} = -\frac{\Sigma S_1 S_2 \Sigma G S_2 - \Sigma G S_1 \Sigma S_2^2}{(\Sigma S_1 S_2)^2 - \Sigma S_1^2 \Sigma S_2^2}$	$\hat{\beta} = \frac{K\Sigma g x - \Sigma g \Sigma x}{K \Sigma x^2 - (\Sigma x)^2}$
Estimate of γ (b)	$\hat{\gamma} = \frac{\Sigma w S_1 S_2 \Sigma w Y S_1 - \Sigma w S_1^2 \Sigma w Y S_2}{(\Sigma w S_1 S_2)^2 - \Sigma w S_1^2 \Sigma w S_2^2}$	$\hat{\gamma} = \frac{\Sigma w S_1 S_2 \Sigma w L S_1 - \Sigma w S_1^2 \Sigma w L S_2}{(\Sigma w S_1 S_2)^2 - \Sigma w S_1^2 \Sigma w S_2^2}$	$\hat{\gamma} = -\frac{\Sigma S_1 S_2 \Sigma G S_2 - \Sigma S_1^2 \Sigma G S_2}{(\Sigma w S_1 S_2)^2 - \Sigma S_1^2 \Sigma S_2^2}$	γ is estimated by an iterative procedure -- see footnote (a).
Estimate of α	$\hat{\alpha} = \frac{\Sigma w y - \hat{\beta} \Sigma w s_1 - \hat{\gamma} \Sigma w s_2}{\Sigma w}$	$\hat{\alpha} = \frac{\Sigma w \ell - \hat{\beta} \Sigma w s_1 - \hat{\gamma} \Sigma w s_2}{\Sigma w}$	$\hat{\alpha} = -\frac{\Sigma g - \hat{\beta} \Sigma s_1 - \hat{\gamma} \Sigma w s_2}{K}$	$\hat{\alpha} = \frac{\Sigma g - \hat{\beta} \Sigma x}{K}$

(a) Weibull's function is fitted using an iterative procedure, i.e., the correlation between g and x is maximized by iterating with regard to γ. This correlation coefficient is given by

$$\rho = \Sigma GX \,/\, \sqrt{G^2 X^2} .$$

(b) β and γ are estimated by adjusting the various variables about their means, i.e.,

$Y = y - \Sigma w y / \Sigma w$; $\quad L = \ell - \Sigma w \ell / \Sigma w$; $\quad G = g - \Sigma g / K$;

$S_1 = s_1 - \Sigma w s_1 / \Sigma w$; $\quad S_2 = s_2 - \Sigma w s_2 / \Sigma w$; $\quad X = x - \Sigma x / K$.

EXISTING DATA [1]. The mean-square error associated with fitting the two- and three-parameter functions appears in Table 4. Although the two-parameter functions are similar, the logistic and the extreme value functions fit the data slightly better than the integrated normal curve. See Figure 5. In turn, the three-parameter functions fit the data somewhat better than the two-parameter functions as shown in Figure 6. However the third parameter is required for only about one-half the data.

Table 5 emphasizes the similarities in the descriptive abilities of these functions. The respective (calculated) 10, 50, and 90 per cent responses are identical for practical purposes. These functions differ only at their tails as indicated by the extrapolated 0.1 per cent response. (These 0.1 per cent responses are computed only for illustrative purposes, and are not intended for use in design.)

Clearly, these data are not adequate to discern which function, if any, precisely describes the nature of the fatigue-limit. Consequently, further experimental study is required. The second part of this paper deals with the design of these tests.

PART II - DESIGN OF FUTURE FATIGUE-LIMIT EXPERIMENTS

EXPERIMENT DESIGN. The design of fatigue-limit experiments must overtly reflect efficiency in terms of over-all cost. Thus it is imperative to exploit fatigue testing. In turn, two considerations are basic to exploitation of fatigue testing:

(1) the minimum number of specimens required (for testing at a given alternating stress amplitude) to attribute a prescribed level of confidence in the position of the datum point, and

(2) preselected spacing of the different alternating stress amplitudes (datum points) to describe the distribution in an efficient manner.

The following discussion shows how simple statistical concepts can be used to design more efficient fatigue tests.

TABLE 4. RESPONSE MEASURED IN TERMS OF MEAN SQUARE ERROR

Test Series No.	Material	Ultimate Strength S_u (ksi)	Two-Parameter Distributions			Three-Parameter Distributions			
			Integrated Normal	Logistic	Extreme Value--Type I	Integrated Normal	Logistic	Extreme Value--Type I	Weibull
(1)	SAE 4330 Unnotched	130	1.145	0.872	0.794	0.571*	0.586	0.822	0.983
(2)	SAE 4340 Unnotched	239	0.511	0.484	1.472	0.766	0.708	0.699	0.789
(3)	SAE 4350 Unnotched	300	2.107	1.308	0.930	0.022***	0.003***	0.491	0.971
(4)	SAE 4350 Notched (1)	300	4.13	3.114	1.525	0.351**	0.307**	0.0001***	3.241
(5)	SAE 4340 Notched	190	0.533	0.132	0.127	0.119*	0.084	0.133	0.136

(1) $K_t = 2.6$

* Level of Significance for γ Pr < 0.05

** " " " " " Pr < 0.01

*** " " " " " Pr < 0.001

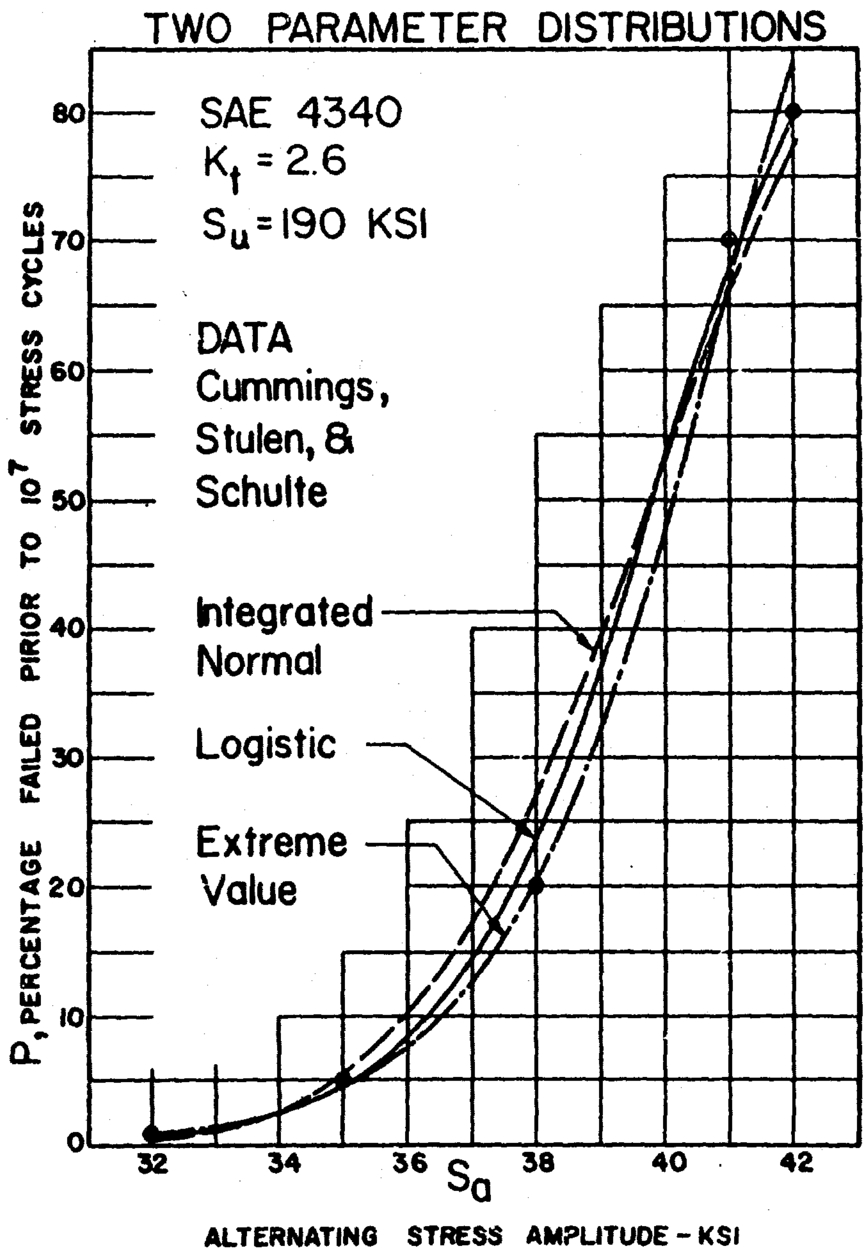


Figure 5 Typical Performance of the Two-Parameter Functions

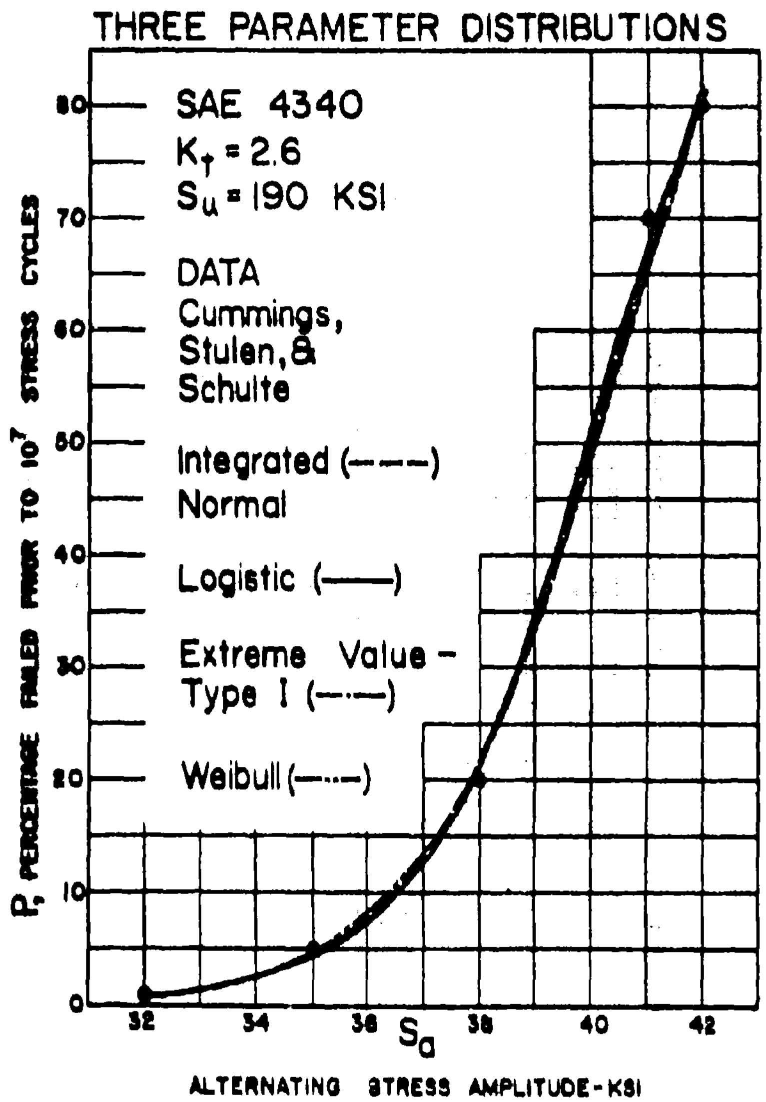


Figure 6 Typical Performance of the Three-Parameter Functions

TABLE 5(a). COMPARISON OF STRESS AMPLITUDES PREDICTED BY VARIOUS DISTRIBUTIONS FOR 10, 50, AND 90 % FAILURES.

The Integrated Normal, Logistic, and Extreme Value--Type I responses pertain to the respective two-parameter functions. The response predicted by the Weibull function is typical of the responses predicted by the other three-parameter functions. (The Weibull γ parameter has no physical meaning.)

	10 % Response				50 % Response				90 % Response			
Material (S_u)	Integrated Normal	Logistic	Extreme Value--Type I	Weibull	Integrated Normal	Logistic	Extreme Value--Type I	Weibull (γ)	Integrated Normal	Logistic	Extreme Value--Type I	Weibull
SAE 4330 (130)	62.61	62.85	62.96	62.72	69.11	69.09	69.76	69.20 (-150)	75.61	75.34	74.10	73.63
SAE 4340 (239)	82.05	82.14	81.73	81.99	89.51	89.53	90.95	89.70 (72.2)	96.98	96.93	96.82	97.02
SAE 4350 (300)	91.87	92.49	92.36	92.07	99.77	99.93	100.49	100.26 (-300)	107.68	107.37	105.67	105.54
SAE 4350* (300)	54.15	54.53	53.86	53.24	60.78	61.03	60.99	61.04 (-300)	67.42	67.53	65.53	65.04
SAE 4340* (190)	35.95	36.34	36.55	36.36	39.80	39.85	40.15	40.09 (0.0)	43.64	43.36	42.45	42.67

* Notched K_t = 2.6

TABLE 5(b). COMPARISON OF THE (Extrapolated) STRESS AMPLITUDES PREDICTED BY THE VARIOUS DISTRIBUTIONS FOR 0.1 % FAILURE.

Test Series	Material	Two-Parameter Distributions			Three-Parameter Distributions
		Integrated Normal	Logistic	Extreme Value--Type I	Weibull*
1	SAE 4330	53.44	49.47	46.13	46.90
2	SAE 4340	71.52	66.28	58.93	74.54
3	SAE 4350	80.73	76.53	72.25	72.65
4	SAE 4350	44.78	40.61	36.24	37.08
5	SAE 4340	30.53	28.81	27.63	28.56

* The other three-parameter distributions are not listed here because: (a) the third parameter is not needed, or when it is needed, (b) the 0.1 % values correspond to "imaginary roots" of the quadratic expression.

The minimum number of fatigue specimens required for testing at a given alternating stress amplitude may be deduced by considering the possible variation in the observed quantal response. For simplicity, assume that the specimen response is described by a binomial distribution that has parameters P and $\sigma_p^2 = \frac{PQ}{N}$ and a coefficient of variation of C. V. = $\sqrt{\frac{Q}{NP}}$. Reliable estimates of P require a small C. V.--on the order of 0.2. Thus, approximately 225 specimens should be tested to estimate P = 0.1. No such experimental results are available. Moreover it is likely that none will be forthcoming in the immediate future because this test alone could cost up to $10,000. (A single fatigue machine running at 10,000 RPM night and day would take eight years to complete such a test if the desired fatigue life is 5×10^8 cycles).

Clearly, statistical efficiency must be sacrificed in fatigue-limit tests. A coefficient of variation on the order of 0.5 is probably the best that can be expected. Even then, approximately 400 specimens are required to estimate P = 0.01. Thus, it appears that the coefficient of variation approach to deducing the number of fatigue specimens required in testing will satisfy neither the statistician nor the materials analyst.

It is possible to mitigate this problem somewhat by estimating the number of specimens required by a different approach, viz., selecting N such that the parameters have a negligible bias. The logistic function is selected here to illustrate this approach. (This selection is made on the basis of ease of computation....there is relatively little difference in the descriptive abilities of any of the functions considered here within the probability ranges of existing fatigue-limit data.)

The linear transform of the logistic function is given by

(1) $$\ell = \alpha + \beta s + \epsilon$$

where ϵ is the (random) error associated with ℓ . This transform is used to fit the logistic function to the data. However, to accomodate subsequent graphical solution of β , this transform is temporarily redefined as [2, 3]

(2) $$\ell' = \ln\left[\frac{P + \frac{1}{2N}}{Q + \frac{1}{2N}}\right] = \alpha + \beta s + \epsilon.$$

The error and variance of ℓ' are given by

(3) $$\pm E(\ell' - \alpha - \beta s) = e^{-Npq}\left\{Npq\ln 3 + \frac{1}{2!}(Npq)^2\ln 5 + \frac{1}{3!}(Npq)^3\ln 7 + \ldots\right\} - \ln(2Npq)$$

(4) $$V(\ell') = e^{-Npq}\left\{Npq(\ln 3)^2 + \frac{1}{2!}(Npq)^2(\ln 5)^2 + \ldots\right\} - e^{-2Npq}\left\{Npq\ln 3 + \frac{1}{2!}(Npq)^2\ln 5 + \ldots\right\}^2$$

and the asymtotic mean and variance are:

(5) $$E(\ell') = \alpha + \beta s$$

(6) $$V(\ell') = \frac{1}{Npq}.$$

Thus, it is clear that bias is a function of Npq. This relationship is shown in Figure 7, where it can be seen that a value of Npq of two or larger affords unbiased estimates of the populations parameters. Accordingly, the minimum number of specimens required at a given alternating stress amplitude can be read from Figure 8.

The spacing of the different alternating stress amplitudes should be sufficiently wide to attain an efficient estimate of β. Considering the logistic function:

(7) $$\hat{\beta} = \frac{\ell_1' - \ell_2'}{s_1 - s_2}$$

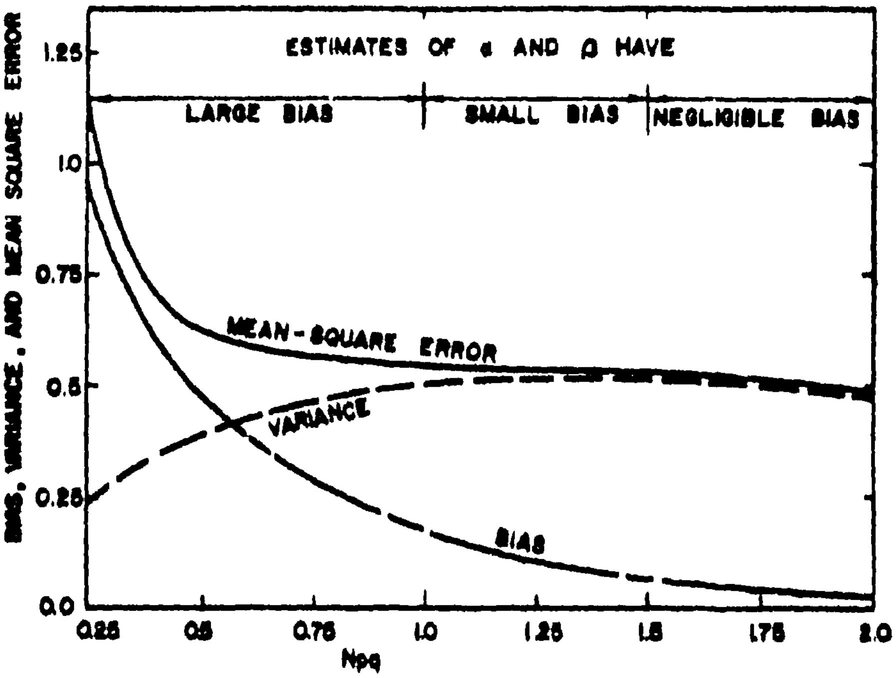


Figure 7 Bias, Variance, and Mean Square Error for $\hat{\lambda}'$

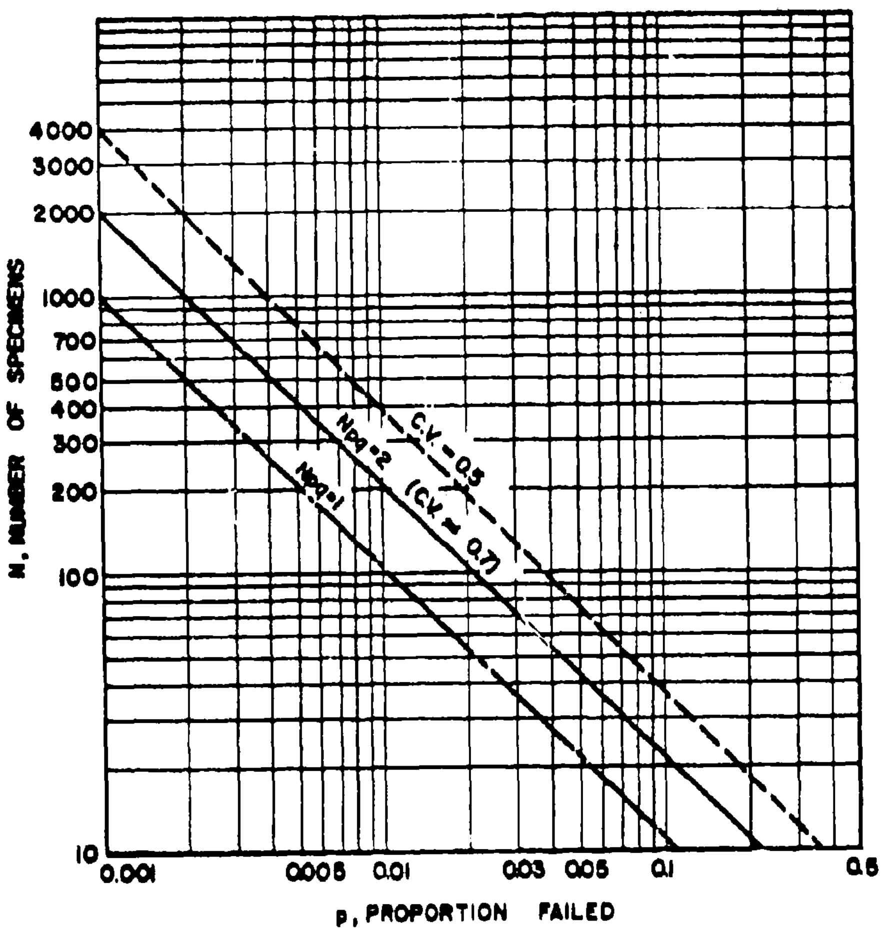


Figure 8 Minimum Number of Specimens Required to Estimate the Population Parameters Efficiently. (C.V. is coefficient of Variation)

where $\ell_1' > \ell_2'$; $s_1 > s_2$; and $(s_1 - s_2) = d$, the spacing.

Equation 7, restated in terms of d, becomes

$$(8) \qquad d = \frac{1}{\hat{\beta}} \left[\ell_1' - \ell_2'\right].$$

Now, selecting p_2 such that

$$(9) \qquad p_1 - p_2 > t \sqrt{\left(\frac{pq}{N}\right)_1 + \left(\frac{pq}{N}\right)_2}$$

this inequality can be rewritten as

$$(10) \qquad p_2 < p_1 - t \sqrt{\left(\frac{pq}{N}\right)_1 + \left(\frac{pq}{N}\right)_2} \quad .$$

Finally, substitution of Equation (10) into Equation (8) gives the desired spacing

$$(11) \qquad d_{min} = \frac{1}{\beta} \left[\ln\left\{\frac{p_1 + \frac{1}{2N_1}}{q_1 + \frac{1}{2N_1}}\right\} - \ln\left\{\frac{p_1 + \frac{1}{2N_2} - t\sqrt{2\,\frac{N_1^2 + N_2^2}{N_1^2 N_2^2}}}{q_1 + \frac{1}{2N_2} - t\sqrt{2\,\frac{N_1^2 + N_2^2}{N_1^2 N_2^2}}}\right\} \right]$$

when $Npq = 2$. This spacing is shown in Figure 9. Note that the spacing can be qualitatively deduced from Equation 9 which indicates that $p_1 - p_2$ can approach zero as N becomes very large.

HYPOTHETICAL FATIGUE TEST. Suppose that the materials analyst has only 100 AISI-1020 annealed steel specimens (Ultimate Strength = 70 ksi), but wishes to obtain the most information concerning the strength distribution. Figure 10 suggests a trial value of the alternating stress amplitude

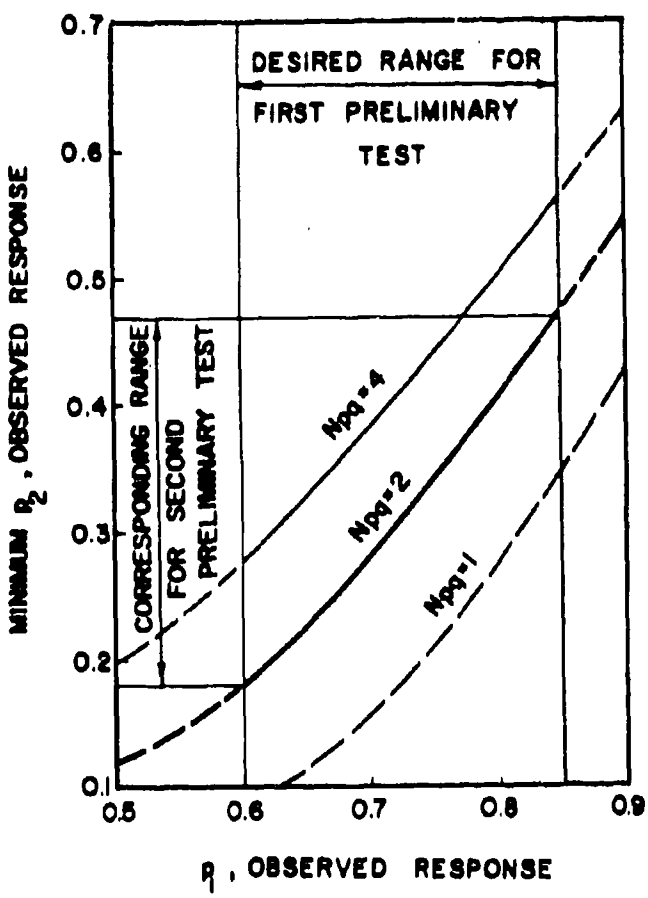


Figure 9a Relationship Between p_1 and p_2 for Efficient Estimation of β.

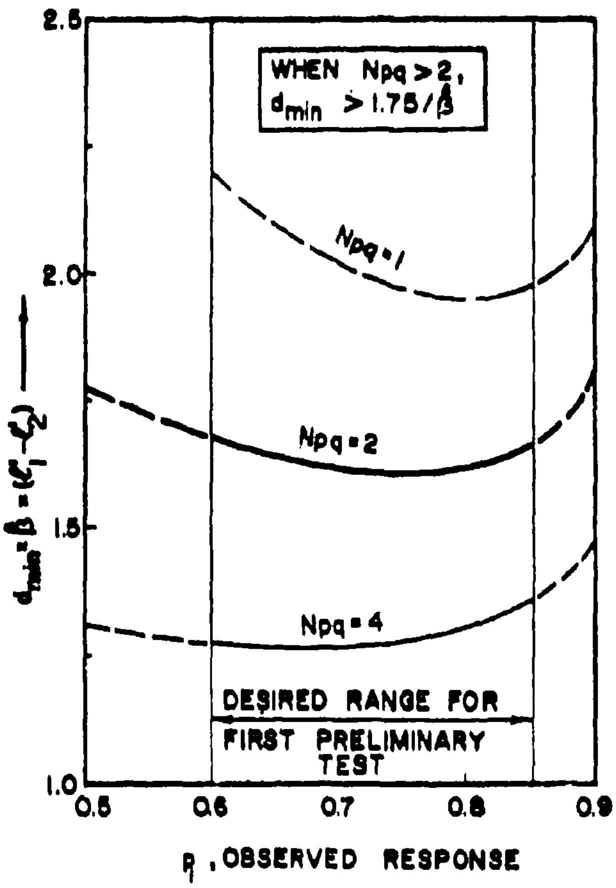


Figure 9b Relationship Between ℓ_1' and ℓ_2' for Efficient Estimation of β. The values of ℓ_1' and ℓ_2' correspond to p_1 and p_2, respectively. Observe that $d_{min} > 1.75/\hat{\beta}$.

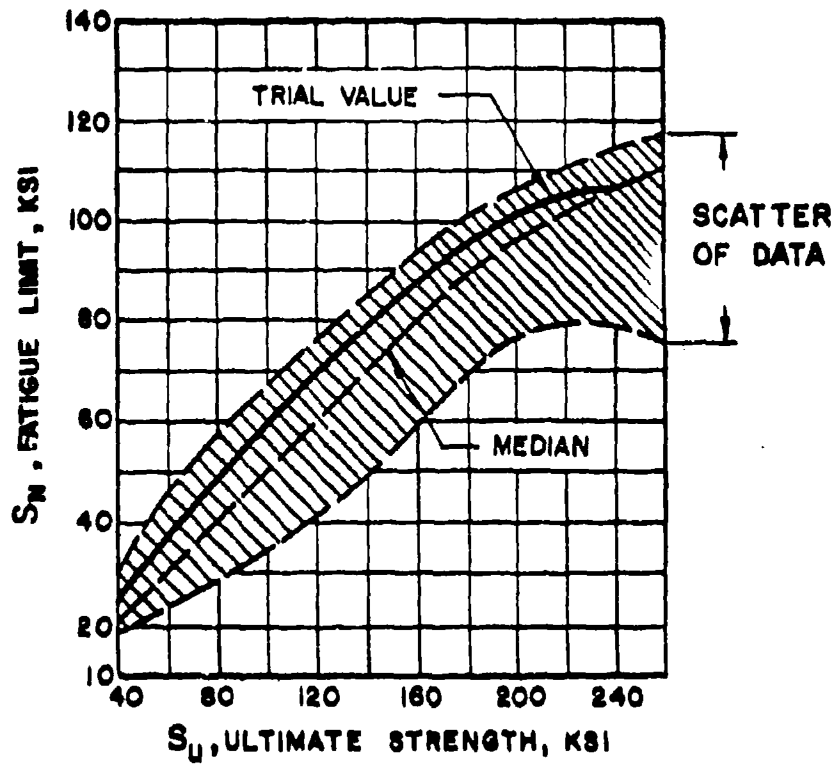


Figure 10 Trial Value for First Preliminary Test Diagram taken from S_N-S_U relationship developed for steel by Bullens [5].

that corresponds to a P of roughly 0.50 to 0.75. Using this trial value, 10 specimens are tested at S_a = 42 ksi and it is observed that 7 specimens fail prior to 10^7 cycles. Setting d = 5 ksi (based on Figure 10), the second test is conducted at S_2 = 37 ksi. In this second test, only 4 of the 10 specimens tested fail. The required spacing in subsequent tests can now be determined by estimating β (using Equation 7). In this hypothetical test

$$\hat{\beta} = \frac{\ell_1' - \ell_2'}{S_1 - S_2} = 0.225 \quad .$$

Thus, d is taken as 7 or 8 and the number of specimens is established as indicated in the following table:

Test	Alternating Stress Amplitude (d=7)	p estimated by graphical solution ($\hat{\beta}$= .225)	Corresponding N for Npq=2	Adjusted N (Npq≈1.75)
Fourth	23	0.04	52	45
Third	30	0.15	16	15
(Second)a	(37)	(0.40)	(10)	(10)
(First)a	(42)	(0.70)	(10)	(10)
Fifth	49	0.90	22	20
		trial total =	110	adjusted total = 100

(a) preliminary tests

Note that Npq is greater than two for the preliminary tests. Actually 10 specimens are not required in either case. The weights can be calculated as these preliminary tests progress and the next test can be started when the respective Npq approaches two. Then, the "saved" specimens can be tested at the most appropriate stress amplitude at the conclusion of the over-all test.

The over-all test data are then listed in tabular form (Table 2) and fitted as outlined in Part I (Tables 1 and 3.)

SUMMARY. Fatigue data will always be somewhat limited because fatigue tests are expensive. Thus, it is necessary to design fatigue tests to be statistically more efficient. This means that care must be given to the preselection of the number of specimens tested and to the spacing of the respective alternating stress amplitudes considered.

Present analyses can only compare the relative performance of different functions with regard to goodness of fit of limited ranges of data.

REFERENCES

1. H. N. Cummings, F. B. Stulen, and W. C. Schulte, "Investigation of Materials Fatigue Problems," Wright Air Development Center, WADC Technical Report 56-611, (March - 1957).

2. F. J. Anscombe, "On Estimating Bionomial Response Relations," Biometrika, 43, 461-4, (1956).

3. S. E. Hitchcock, "A Note on the Estimation of the Parameters of the Logistic Function, Using the Minimum Logit χ^2 Method," Biometrika, 49, 250-252, (1962).

4. D. J. Finney, "Probit Analysis," Cambridge University Press, London, (1952).

5. D. K. Bullens, Steel and Its Heat Treatment, Vol. 1, P. 157, Fifth Edition, J. Wiley & Sons, New York, (1948).

PERTINENT REFERENCES

6. "A Tentative Guide for Fatigue Testing and the Statistical Analysis of Fatigue Data," American Society for Testing Materials, Special Technical Publication No. 91-A. (Supplement to Manual on Fatigue Testing, ASTM, STP, No. 91 (1958).

7. E. J. Gumbel, "Statistics of Extremes," Columbia University Press, New York, (1958).

8. A. M. Freudenthal, and E. J. Gumbel, "On Statistical Interpretation of Fatigue Tests," Proc. Roy. Soc., A, 216, 309, (1953).

9. W. Weibull, "A Statistical Representation of Fatigue Failures in Solids," Trans. Roy. Inst. Tech. Stockholm No. 27 (1949).

10. A. H. Soni and R. E. Litt' , "Statistical Analysis of Fatigue Limits Using the Logistic Function" The University of Michigan, Industry Program of the College of Engineering, IP-634, (Oct. 1963) Also see, "Materials Research and Standards," Vol. 4, No. 9, pp. 471-473, September 1964.

GETTING REGRESSION ANALYSIS IMPLEMENTED*

W. H. Ammann
U. S. Army Aviation Materiel Command
St. Louis, Missouri

INTRODUCTION. The idea for this presentation came as a result of unsuccessful attempts to solve an analytical problem which was complicated by restraints placed on the collection of data for analysis. Figure 1. This situation is not an isolated one but generally occurs when much data are already being gathered and they are not sufficient for the analysis desired. Alteration of the existing data collection system just to satisfy the needs of a supposedly isolated and parochial study effort is generally not feasible. So, it is necessary to consider the existing data limitations as part of the problem to be solved.

In this case, the success of the analytical effort depends on the relationship which is established between the kind and amount of information which is needed to define the problem and the kind and amount of information available for solving the problem as defined. When this problem-defining and solving effort does not provide meaningful results (Figure 2), three questions are appropriate: has the problem been inadequately defined because of ignorance about the nature of the operation being considered?; are the data collected not sufficient in kind and/or quantity to establish the desired relationships?; and are the data being inadequately analyzed because of the ignorance of the analysts? It is generally necessary to assume that data collected for analysis are not erroneous to the extent that they would be the principal cause for the lack of meaningful analytic results because it is seldom feasible to double check the correctness of the data.

ONE EXAMPLE. To illustrate the foregoing remarks, a recent study will now be described. To appreciate the need for this study, it is necessary to point out that AVCOM's supply effort relates to keeping Army aircraft from being too often deadlined due to a lack of parts (commonly referred to as an Equipment Deadlined for Parts or briefly an EDP situation) while incurring no more than the least costs necessary to obtain such results. It was recognized that this effort might be made more

*The views contained herein have not been approved by the Department of the Army, and represent only the views of the author.

effective and/or efficient if it could be analytically demonstrated how the rate, at which aircraft are EDP, varies with various supply actions and ultimately with the costs associated with each action.

A study to obtain the desired analytical results was developed.

a. Concept: It was recognized that the total time that aircraft are EDP is affected by how often an EDP situation occurs and how long it takes to satisfy each EDP situation. Therefore, the study was naturally subdivided into a study of the frequency of occurrence of EDP situations and a study of the time required to satisfy EDP situations. Immediately, obstacles were encountered.

(1) Frequency of Occurrence: How often each aircraft is EDP during the month is not reported. However, when an aircraft is EDP for a part which is supposed to be furnished by AVCOM action and that part cannot be obtained below depot level, an EDP requisition is sent to AVCOM. Therefore, there are at least as many instances of aircraft EDP as there are valid EDP requisitions received at AVOCM and it is operationally known that there are more such occurrences since aircraft are EDP for parts which are supplied without AVCOM action. The term valid EDP requisitions is needed because those for common hardware parts which could not render an aircraft operationally deadlined were excluded as being invalid.

Fortunately, the total time that each aircraft is EDP is reported to AVCOM. So, it was hoped that an estimate of the amount of change which might be achieved in the days aircraft are EDP by a supply action which might reduce the rate at which EDP requisitions occur at AVCOM by a particular amount, might be obtained by regression analyses by aircraft type. The results of these analyses will be indicated later.

(2) Time to Satisfy: Meanwhile, attempts to relate the time to satisfy an aircraft EDP situation encountered similar data constraints. When an aircraft EDP situation is satisfied without an AVCOM action in response to an EDP requisition, the time required to obtain such satisfaction is unknown at AVCOM. So, it was necessary to assume that such instances have a random effect on the total time aircraft are EDP in a month. Then, a meaningful correlation might be discovered between the time aircraft are EDP and the time required to satisfy an EDP requisition at AVCOM.

Also, the complete time required to satisfy an EDP requisition at AVCOM could not be easily obtained. The time that was obtained is the time between the date an EDP requisition is initiated and the date on which materiel release at the appropriate depot is confirmed.

In other words, the time consumed after a material release confirmation is sent to AVCOM and until the part arrives at the site of the particular EDP aircraft was not readily measurable and had to be left out of the study. Again, it was necessary to assume that the effect of this time or the total time aircraft are EDP is random and that a meaningful correlation might exist between the time aircraft are EDP and the principal portion of the time to satisfy an EDP requisition measured in this study.

On the other hand, since an EDP requisition does not identify the specific aircraft which is awaiting the part, it is possible that an EDP aircraft has been made serviceable by using a part obtained from some other source such as off of a crash damaged aircraft and yet the pertinent EDP requisition is not satisfied. It was hopefully assumed that such instances might compensate for some of the excluded shipping time.

b. Sample Selection: By now, hopes to obtain fruitful analytical results were waning and yet the worse was yet to come. Since it was desired to obtain some useful results as soon as possible and the information about aircraft days EDP is available only on a monthly basis, six months data or six data points were chosen for analysis. After the data were gathered, there was reason to believe that data concerning all EDP requisitions received by AVCOM during the first three months observed had not been obtained. Further, it could not be determined whether the sample EDP requisitions could be validly claimed to be a representative sample. Therefore, only the latter three month's data were used for regression analysis. At this point, the problem being described can be summarized as shown. Figure 3.

c. Results Obtained: Approximately nine months elapsed before efforts to obtain the preferred analytical results were exhausted. A total of 14 aircraft types were considered. Needless to say, the results obtained were disheartening even though not unexpected.

(1) Frequency of Occurrence: Table 1 contains estimates of the relationship between the days aircraft are EDP and the quantity of EDP requisitions received at AVCOM.

(2) Time to Satisfy: Table 2 contains estimates of the relationship between the days aircraft are EDP and the major portion of the time to satisfy EDP requisitions at AVCOM.

(3) It is recognized that three data points are not enough to preclude apparently conflicting results from occurring because of sampling variations but there were no more reliable data points which could be used to reduce this likelihood. However, the occurrence of both positive and negative correlation coefficients is disconcerting. In the case of negative ones, it is implied that a reduction in aircraft days EDP can be obtained by increasing the frequency of occurrence of EDP instances or by taking more time to satisfy EDP requisitions. Both of these implications are unreasonable. With the hope that the three questionable data points might be good ones, regression analyses using six points were made but no more reasonable results were obtained.

(4) To preclude some wrong implications, it must be pointed out that this nine month study effort did not consume much more than one analyst's time. The study time had to take six months to obtain six months of data. Additional time was required to allow EDP requisitions received near the end of the sixth month to be satisfied. In addition, several by-product analyses were made with the data collected. In other words, it would be unfair to conclude that this analytical effort was not worthwhile. Also, it seems that it could be concluded that the desired results were not obtained for at least the first two of the reasons listed on Figure 2; namely, inadequate representation of the problem and insufficient data collected both in type and quantity.

d. Question: However, the question still remains: What can be done to increase the effectiveness of the analytical effort being expended in the manner just described?

ANOTHER EXAMPLE. Before attempting to present any subjective answers to the question just stated, another analytical problem area can be used to suggest that there is a related question that also needs answering. This analytical problem is suggested by a review of budgeting and funding practices.

It is not necessary to know the exact budgeting and funding procedures to appreciate the features which are useful here. Figure 4.

a. The preparation of a budget must be in accordance with guidance furnished by higher headquarters. This guidance has usually been different from year to year. This situation implies that a generally sound budgeting procedure has not yet been determined.

b. In addition, forecasted budget requirements are never completely honored. Somewhere up the line, limitations are set below the accumulated forecasted requirements and these limitations are somehow partitioned and passed down to each organization involved.

c. Further, each organization's general objective is to make commitments nearly equal the limitations appropriate at the time of each within year review. In other words, if there is only a mid-year review, commitments should be nearly equal to one half of the annual limitation otherwise it might be concluded that even less funds will suffice and limitations will be decreased accordingly. As a result of these within year reviews, particular fund limitations for the remainder of the year are revised; sometimes upward and sometimes downward.

d. At this point, it is well to hypothesize the logic which supports this budgeting and funding practice. It is initially assumed that no one can forecast an organization's budgetary requirements more accurately than the organization itself. Therefore, forecasted requirements are made by each organization and these are the starting point for the budgeting cycle. Since fund limitations have always been set less than forecasted budget requirements, organizations find it expedient to compensate for such reduction by somehow inflating estimates of requirements. It seems reasonable to assume that the extent of this inflation cannot be accurately determined by the people who set limitations otherwise budgetary guidance could preclude such inflation and forecasted requirements could be honored. Also, since the practice of setting limitations less than forecasted requirements has never been considered responsible for serious operational shortages, the practice has been continued without fear.

It seems that this strategic exercise must persist until it has been definitely learned that the allotment of different quantities of funds leads to the achievement of recognizably different accomplishments. Only then can superiors choose the desired amount of accomplishments and fund accordingly. Thus, the question arises:

How can the regression analysis effort, necessary to establish a sound relationship between money allotted and results achieved thereby, be obtained?

CONCLUDING REMARKS. In review, it seems that the two situations just described indicate a need for a way; of improving the effectiveness and efficiency of the analytical effort trying to do regression analysis in a subordinate command such as AVCOM; and of getting regression analysis implemented in a higher headquarters in the budgeting-funding subject area.

a. In the first case, it is possible to take the viewpoint that certain analytical efforts should be dropped when data collection restraints are too restrictive or that it is worthwhile to expend some effort to remove as many of those restraints as necessary. However, the potential value of certain analytical results is sufficient to preclude their being dropped until it has been indisputably demonstrated that they cannot be obtained in spite of existing constraints. Also, the removal of data collection restraints to satisfy local analytical needs is practically impossible since existing data collection and reporting requirements have been entrenced by tradition and austerity measures in the manpower area preclude the collection and reporting of additional data for local analyses that have not been specifically required by higher headquarters. Therefore, it seems that some outside, authoritative intervention is needed if the situation confronting local, investigative analyses is to be improved.

b. In the second case, since budgeting guidance is furnished by higher headquarters and must be adhered to by subordinate commands, it seems that regression analysis must be attempted and found successful at the top before the official authorization to do such analysis at the bottom can be expected and before the cooperation necessary to have a reasonable chance at being successful with this effort will be forthcoming.

In other words, it seems that it is not enough to hire analysts at all levels in the Department of the Army and then allow organizational tradition to render such analysts ineffective and inefficient. The situation could be significantly improved if (Figure 5) the Office of the Chief of Research and Development (OCRD) would form a Survey Team of renowned analysts who would visit selected Army headquarters to determine the extent and kind of analytical program that seems appropriate for each organizational

level and would then prepare a recommended Department of the Army program. Then, OCRD could coordinate this program as appropriate and direct that the coordinated program be done. This type of positive action seems a bit extreme and probably impossible to obtain and so a solicitation for a less extreme improvement action and one more within the authority of a subordinate organization is hereby extended.

Table 1

A/C Type	A/C Days EDP vs Qty of EDP Rqns	Correlation Coefficients
OH-13	y = 1921 + 0.777x	0.927
UH-19	y = 397 + 2.610x	0.903
CH-21	y = -419 +21.177x	0.933
OH-23	y = 2701 - 0.890x	-0.120
CH-34	y = 399 + 5.472x	0.621
CH-37	y = 588 - 5.447x	-0.683
UH-1	y = 1572 + 1.679x	0.415
O-1	y = 519 +20.107x	0.783
U-6	y = 908 - 1.543x	-0.347
U-8	y = 533 - 2.403x	-0.713
U-1	y = 391 - 1.582x	-0.281
OV-1	y = 468 + 2.610x	0.976
CV-2	y = 362 - 0.522x	-0.649
CH-47	y = 885 - 5.567x	-0.997

y = is in terms of aircraft days EDP

x = is in terms of quantity of EDP requisitions received at AVCOM

Table 2

A/C Type	A/C Days EDP vs Qty of EDP Rqns	Correlation Coefficients
OH-13	y = 1923 + 0.126x	0.998
UH-19	y = 567 + 0.103x	0.394
CH-21	y = 509 - 0.046x	-0.088
OH-23	y = 3194 - 0.457x	-0.662
CH-34	y = 482 + 0.382x	0.777
CH-37	y = 577 - 0.220x	-0.883
UH-1	y = 6913 - 1.551x	-0.388
O-1	y = 3151 - 6.119x	-0.498
U-6	y = 950 - .184x	-0.465
U-8	y = 56 + 0.530x	0.999
U-1	y = 302 + .116x	0.681
OV-1	y = 431 + 0.132x	0.619
CV-2	y = 359 - 0.023x	-0.728
CH-47	y = 690 - 0.092x	-0.566

y = is in terms of aircraft days EDP

x = is in terms of the principal quantity of days required to satisfy EDP requisitions received at AVCOM

Problem To Be Solved

Analytical
Problem

Plus

Data Collection
Restraints

FIGURE 1

No Useful Solution Obtained

Inadequate Representation of Problem?

Insufficient Data Collected?

Inadequate Analysis?

FIGURE 2

Sample Problem

Effect of Supply Actions on
Acft EDP Rate

Plus

Inexact EDP Frequency

Unknown Extent of EDP Change
Due to Non-AVCOM Action

Unknown Shipping Times

Only Three Reliable Data Points

FIGURE 3

Budgeting & Funding

Guidance Changes Annually

Limitations Less Than
Forecasted Requirements

Commit Full Limitations

Within Year Reviews

Revised Limitations

FIGURE 4

For Consideration

PROFESSIONAL TEAM:

Conduct Survey
&
Describe Analytical Program

CHIEF, RESEARCH & DEVELOPMENT:

Require Program Be Done

FIGURE 5

ASSESSMENT AND CORRECTION OF DEFICIENCIES IN PERT

H. O. Hartley and A. W. Wortham
Institute of Statistics
Texas A&M University
College Station, Texas

1. INTRODUCTION. As is well known, the technique known under the name of PERT (Program Evaluation Review Technique) is concerned with a 'project' comprising a large number of successive 'activities' which are arranged in a complex 'network' (see e.g. Figure 2). Each activity 'commences' at a particular 'point' of the network but not until all activities 'terminating' at that point are completed. Specifically, PERT is concerned with computing the expected time required to complete all activities of the project: -Assuming that the time taken to complete a particular activity follows a specified distribution of completion times, the total time needed to complete the project the so called 'critical time' is a statistical variable and is given by the total of completion times along the 'critical path', i.e. along that sequence of activities in the network which for a given sample of completion times takes longest to reach every point along its path. The expected value of this critical time is the expected time to complete the project.

Now it is well known that PERT does not compute the correct critical time as defined above but instead uses for each activity the average completion time and then determines a unique and fixed critical path as the sequence of activities for which the sum of the expected completion times is at a maximum. Critical path determination by this method may be badly misleading and may result in a serious underestimate of the expected time to complete the project. Moreover, it may also lead to erroneous information on the identification of 'critical activities', i.e., activities which are crucially responsible for the delay in completion of the project.

Whilst this shortcoming of PERT has been known from its initiation and the above method is deliberately used as an approximate short-cut, we do not think that the magnitude of the bias in this short-cut method is fully appreciated. Indeed it can be shown (see e.g. section 8) that under certain circumstances PERT may underestimate the correct expected completion time by 50% or more. Moreover, for a general network, PERT provides the correct answer only under the (completely unrealistic) assumption that there is essentially no variability in the completion times for each activity.

One of the objectives of this paper is therefore to eliminate this bias from PERT; in fact, we shall provide a method of computing the probability distribution of critical times and thereby supply not only the correct value of its expectation but likewise of its variance and percentage points.

It may rightly be argued that our exact method of critical path analysis is based on the assumed distribution of completion times for each activity, and that there is usually a notorious lack of information on such timings. This point is well taken. However, we feel that the deplorable lack of input data should not excuse us from using a method accurately utilizing at least all the available information. Moreover and more positively our method enables the analyst who is uncertain about the completion times of (say) a particular activity in the network to evaluate the effect of altering his assumptions about that activity on the critical time and path. We consider the provision of such a 'sensitivity analysis of PERT' as an important contribution to planning a project 'under uncertainty'.

Mathematically our method utilizes the following devices: -

(a) A classification of networks into different types depending on their degree of involvement and complexity.

(b) An operational calculus by which the distribution of critical times will be derived by numerical analysis, notably numerical integration. This method will provide the solution to our problem for the basic types of networks.

(c) A Monte Carlo procedure providing an approximate solution for the more involved networks.

(d) Analytic solutions for particularly simple networks and particularly simple distributions of completion times. These are mainly used for illustration purposes.

2. GENERAL DEFINITIONS AND 'UNCROSSED NETWORKS'. In order to provide a mathematically rigorous theory of PERT analysis for networks, it is necessary to introduce certain definitions and concepts. We therefore give the following definitions and explanations: -

2.1. An activity is represented by one or two line segments in the network (see Figure 1). It 'commences' at one of its ringed end points and 'terminates' at the other ringed end point, the 'direction of the time flow being indicated by the arrow. The numbering of the activities is explained in 2.3.

2.2. A Network Point: - These are represented by ringed points in Figure 1. A network point represents any stage in the network occurring at the beginning and/or end of an (or several) activity(ies) (e.g., event 5 in Figure 1 is a network point since activity (7; 2, 5) terminates and activities (10; 5, 8) and (11; 5, 8) commence at that stage of the network.

2.3. Codes: - 'Network points' carry a serial number (ringed in Figure 1) identifying them. The order of the numbering is immaterial at this stage. An activity also carries a 'serial number' (preceding the ;) but also the number of the network point at which it commences followed by the network point number at which it terminates. Thus (7; 2, 5) denotes activity No. 7 commencing at point No. 2 and terminating at point No. 5.

2.4. Two consecutive activities are defined as activities numbered (t; i, j) and (s; j, k) i.e., the first terminates at point j whilst the second commences at point j.

2.5. A Path from i to j is a 'sequence of consecutive activities' starting at point i and finishing at point j (e.g., activities (2; 0, 2), (7; 2, 5) and (10; 5, 8) starting at point (0) and terminating at point (8)).

2.6. A complete path - A path starting at the beginning and finishing at the end of the project (e.g., the path formed by (1; 0, 1), (5; 1, 4), (9; 4, 7) and (15; 7, 10)).

2.7. A Universal Point - A network point through which all complete paths pass (the only universal points in Figure 1 are at 0 and 10).

2.8. Consecutive Points - Point j is consecutive to Point i if both j and i are universal and if all paths starting from i pass through j before passing through any other universal point (if any).

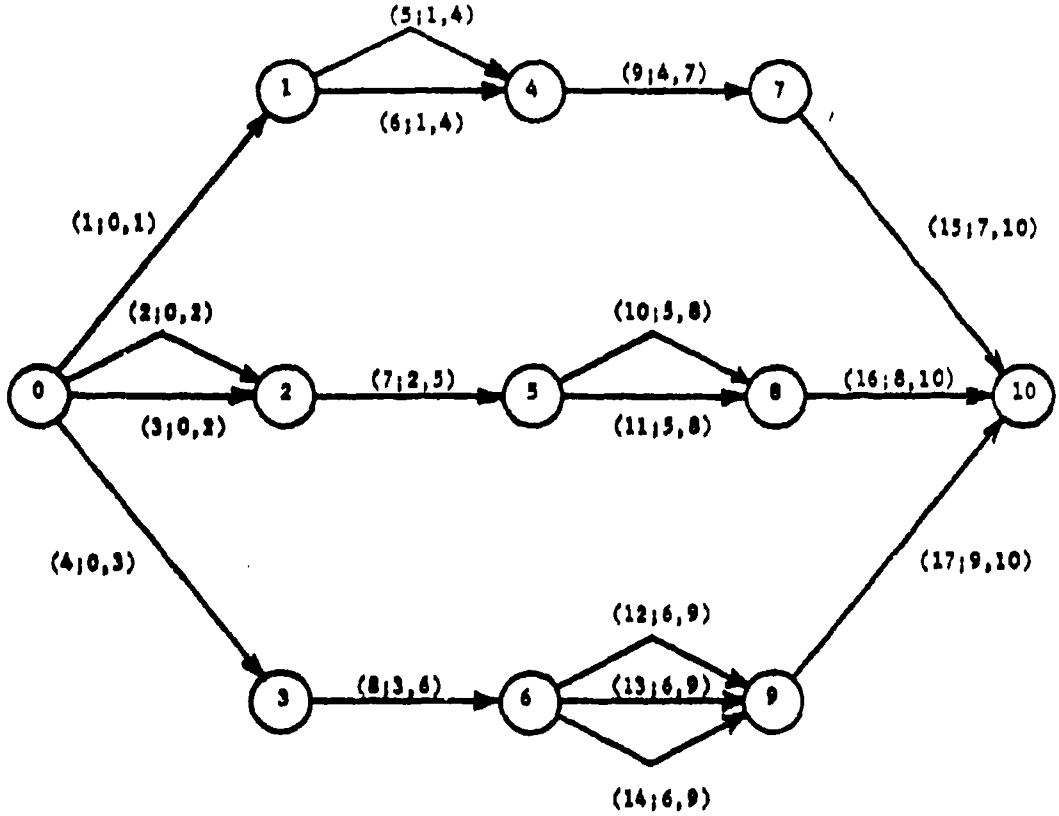


FIGURE 1
NETWORK NOTATION

2.9. Sets of first order branches - Consider the set of all paths commencing at a universal point i and terminating at a universal point j consecutive to i. Subdivide the set of these paths into exhaustive subsets such that any two paths in different subsets have only points i and j in common but any two paths in the same subset have at least one more point in common. (This is always possible since we may place, if necessary, all paths in the same subset.) These mutually exclusive subsets are called '1st order branches.' (e.g., in Figure 1 the paths formed by connecting points 0,1,4,7,10 form the first 1st order branch, the paths formed by connecting points 0,2,5,8,10 the second 1st order branch and the paths formed by 0,3,6,9,10 the third 1st order branch.) If there are only two consecutive points in the network (i.e., the start and the end) and there is only <u>one</u> set of paths as described above, we shall term it a zero order branch. For example, Figure 3 would constitute a single zero order branch, so would a single activity network.

2.10. Sets of 2nd order branches - Consider a particular 1st order branch starting at a universal point i and ending at a universal point j consecutive to i and regard it as a separate network. Apply definitions 2.1 to 2.9 to this network, then any 1st order branches of this first order branch are called second order branches, but any zero order branch of a first order branch still be called a 1st order branch. (e.g., in Figure 1 activities (2; 0, 2), and (3; 0, 2) connecting points (0) and (2) are two second order branches belonging to the second first order branch. Likewise (7; 2, 5) is a second order branch belonging to this first order branch.

2.11. The uncrossed network - If by the repeated application of definitions 2.1 to 2.10 all individual activities in the network can be identified as different k-th order branches (for some $k \geq 0$), the network is said to be "uncrossed." (e.g., the network in Figure 1 is uncrossed and all activities are recognized as different 2nd order branches. The network in Figure 2 is likewise uncrossed with some of the individual activities being 2nd order branches and some 3rd order branches. However, the network in Figure 3 is crossed - there being only one (0 order) branch comprising all activities.

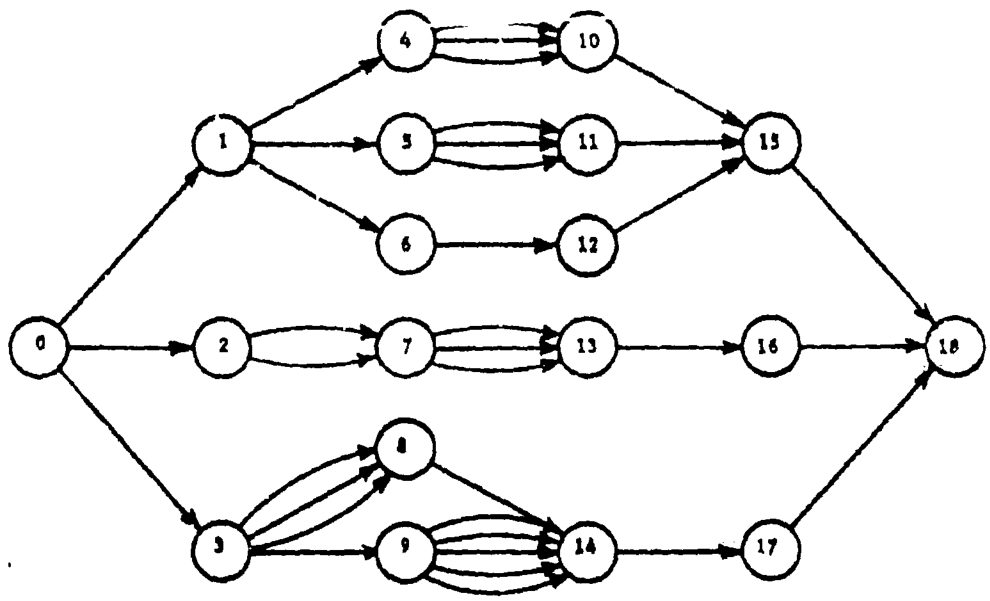


FIGURE 2
AN EXAMPLE OF AN "UNCROSSED" NETWORK

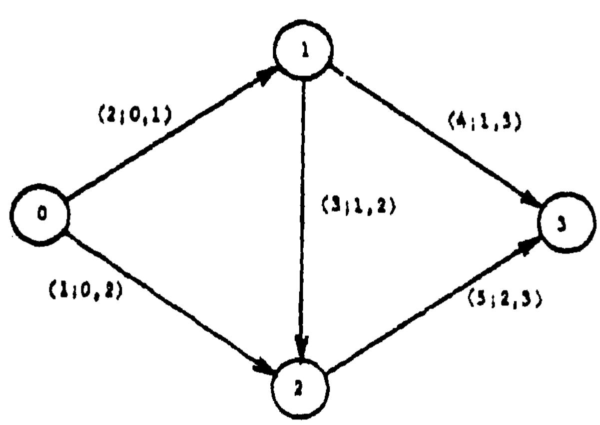


FIGURE 3
AN EXAMPLE OF A "CROSSED" NETWORK
"WHEATSTONE BRIDGE"

3. CROSSED AND MULTIPLE-CROSSED NETWORKS. The arrangement shown in Figure 3, called the 'Wheatstone bridge', has been quoted in the previous section as an example of a crossed network. It consists of the five activities (1; 0, 2), (2; 0, 1), (3; 1, 2), (4; 1, 3) and (5; 2, 3). If now each of these five single activities is replaced by an uncrossed network, as defined in Section 2, we shall reach a network called a '1st order crossed network.' More specifically we define a 0-order crossed network as an uncrossed network in which at least one of the 'activities' is replaced by a Wheatstone bridge (see Figure 3). With the help of this network we define a t^{th}-order crossed network (for $t \geq 1$) as a 0-order crossed network in which any 'activity' may be replaced by a k^{th}-order crossed network with $0 \leq k \leq t-1$, but at least one activity is replaced by a $(t-1)^{st}$-order crossed network.

Although most practical situations of activity networks will be recognized as a t^{th} order crossed network for some order t. There are clearly quite small networks which do not belong to this category, as for example the network shown in Figure 4:

4. OPERATORS FOR EXACT SOLUTION BY NUMERICAL ANALYSIS. Consider first the case of an uncrossed network as defined in section 2. It is easy to show (see e. g. Section 5) that an uncrossed network can be built up from individual activities by two basic operations which can be briefly described as follows: -

Operation π: - Placing activities in parallel

Operation S: - Placing activities in series

These basic operations, well known concepts in electric circuit theory, are illustrated in Figures 5 and 6.

Corresponding to these two basic networks we now develop the simple equation for the c. d. f. (cumulative distribution function) of the 'critical time' in the two basic networks.

a. Parallel activities: -

Denote the serial number of the k activities in parallel by s so that $s = 1, 2, \ldots, k$ ($k = 5$ in Figure 5) and denote the time required to

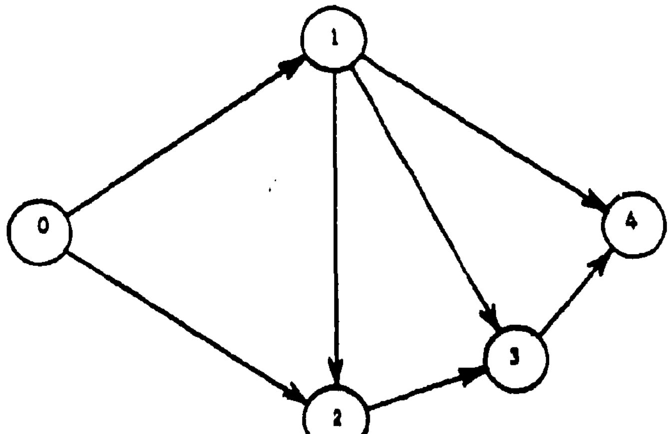


FIGURE 4
EXAMPLE OF A NETWORK NOT IDENTIFIABLE
AS t^{th} ORDER CROSSED NETWORK

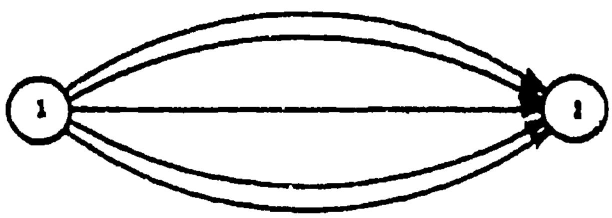


FIGURE 5
ACTIVITIES IN PARALLEL

FIGURE 6
TWO ACTIVITIES IN SERIES

complete the s-th activity by t_s. If the c.d.f. of t_s is denoted by $F_s(t_s)$ then the critical time t for this simple network is clearly given by $t = \max_s t_s$ so that the c.d.f. of t is obtained as

$$F(t) = \Pr\left\{\max t_s \le t\right\} = \prod_{s=1}^{k} F_s(t) \tag{1}$$

b. Two activities in series.

Denote the times required to complete the two activities by t_1 and t_2 respectively and their c.d.f.'s by $F_1(t_1)$ and $F_2(t_2)$. Then the critical time for this simple network is clearly given by $t = t_1 + t_2$ so that the c.d.f. of t is obtained as

$$F(t) = \int_0^F F_1(t-t_2)dF_2\ ,\ \text{where } F = F_2(t) \text{ and } t_2 = F_2^{-1}(F_2). \tag{2}$$

It should be noted that equations (1) and (2) yield the c.d.f. F(t) for the basic network from the c.d.f.'s of the individual activities. Therefore, these basic networks can subsequently be regarded as 'individual activities' and entered as $F_s(t_s)$ in subsequent operations of the type (1) and (2). It is obvious therefore that by repeated application of (1) and (2) the c.d.f. of an uncrossed network such as in Figure 1 and Figure 2 can be obtained. The operational logic for this is given in section 5.

Next we deal with 1st order crossed networks and to this end must evaluate the c.d.f. of the critical time t for the Wheatstone bridge (figure 3). Denoting by $t_1, \ldots, t_5$ the completion times for the five activities $s=1,2,\ldots,5$ as arranged in Figure 3 and by $F_i(t_i)$ their respective c.d.f.'s we obtain by elementary probability calculus the c.d.f. of the critical time t as a sum of three integrals as shown in (3) below: -

$$F(t) = \int^{a} dF_2 \int^{b} dF_3 \int^{c} dF_5 \; F_1(t_3+t_2)\, F_4(t_3+t_5)$$

$$(3) \qquad + \int^{a} dF_2 \int^{d} dF_4 \int^{f} dF_5 \; F_1(t_2+t_4-t_5)\, F_3(t_4-t_5)$$

$$+ \int^{g} dF_1 \int^{h} dF_5 \int^{j} dF_2 \; F_3(t_1-t_2)\, F_4(t_1+t_5-t_2)$$

where $t_i = F_i^{-1}(F_i)$ are the inverse functions of $F_i(t_i)$, all variables of integration are the F_i with integrations starting at $F_i = 0$ and ending at points 'a' to 'j' given by

$$a = F_2(t),\; b = F_3(t-t_2),\; c = F_5(t-t_2-t_3)$$

$$(4) \qquad d = F_4(t-t_2),\; f = F_5(t_4),\; g = F_1(t)$$

$$h = F_5(t-t_1),\; j = F_2(t_1).$$

It should be noted that the three terms in (3) correspond to the three mutually exclusive and exhaustive situations (a), (b), (c) shown below

(a) Critical path $t = t_2 + t_3 + t_5$

(b) Critical path $t = t_2 + t_4$

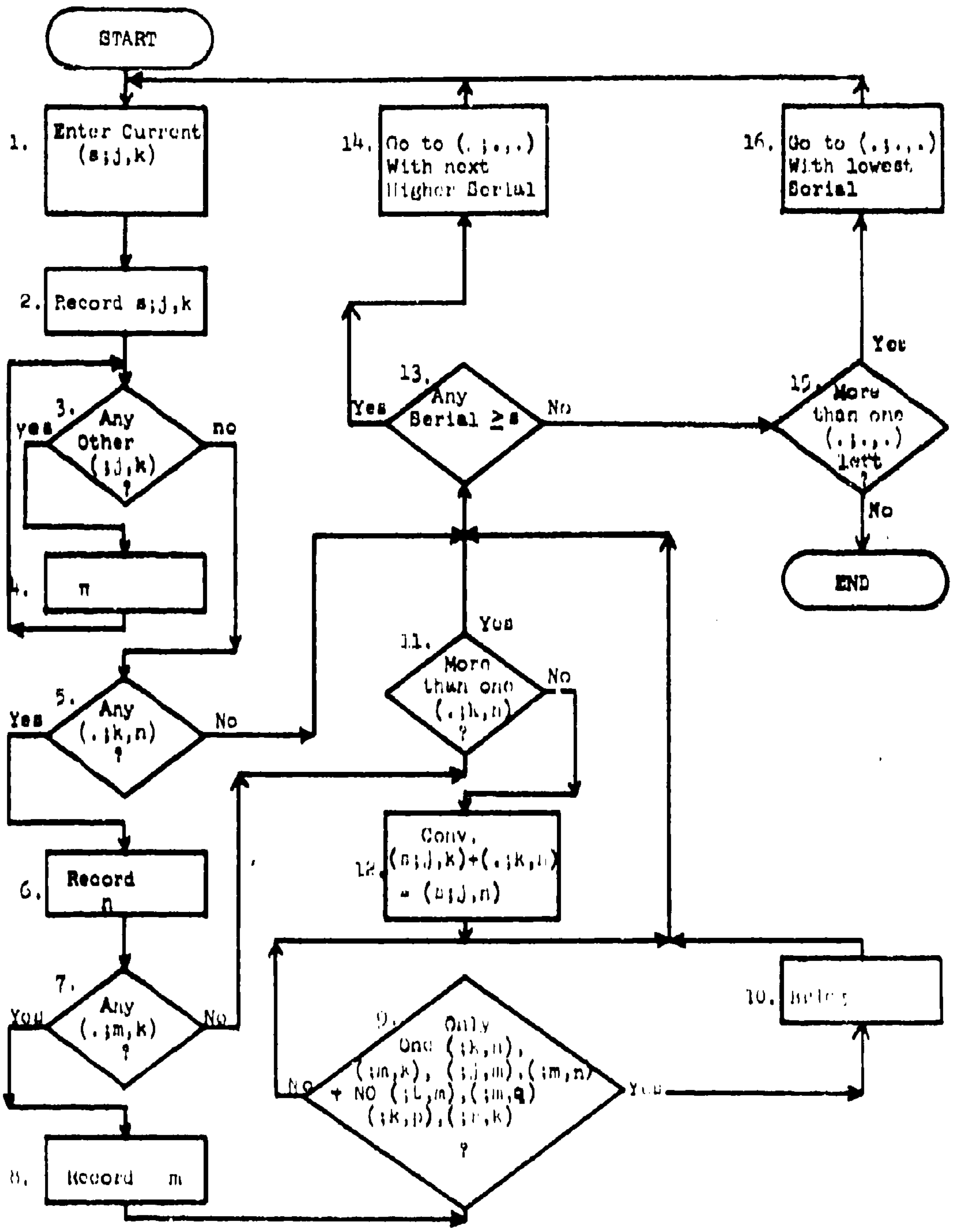
(c) Critical path $t = t_1 + t_5$.

The general case of a t-th order crossed network is finally covered by repeated application of the above operators as shown in section 5.

5. <u>THE COMPUTATIONAL LOGIC FOR t-th ORDER CROSSED NETWORKS.</u> The computer logic shown in Figure 7 will compute the c.d.f. of the critical time in a t-th order crossed network from c.d.f.'s of the completion times of the individual activities.

FIGURE 7

FLOW DIAGRAM FOR t^{th} ORDER CROSSED NETWORKS



The initialization of the computation consists of loading the code numbers of all activities $(s; j, k)$ (see section 2.3) as well as readying the tape giving all their c.d.f. functions. If the serial number of the activity is immaterial we shall use the symbol $(.; j, k)$. In the course of the operations certain code numbers will be deleted and the retained code number activities have their c.d.f.'s modified. We should give the following explanations of some of the operations involved: -

Box 1; 2 An activity $(s; j, k)$ with the current serial number s and starting at j and ending at k (see 2.3) is processed i.e. s; j k are recorded and the associated c.d.f. $F_1(t)$ loaded.

Box 3 A test is made as to whether there is a 2nd activity starting at j and ending at k

Box 4 If the 2nd activity starting at j and ending at k has a code $(u; j, k)$ and a c.d.f. of $F_2(t)$, replace $F_1(t)$ by $F_1(t)\ F_2(t)$ and delete $(u; j, k)$ from the list of code numbers and $F_2(t)$ from the tape of c.d.f. functions.

Box 12 If the c.d.f. functions of activities $(s; j, k)$ and $(\cdot; k, n)$ are denoted by $F_1(t)$ and $F_2(t)$ respectively we replace $F_1(t)$ by $\int_0^F F_1(t-t_2)\ dF_2$ with $F = F_2(t)$, and $t_2 = F_2^{-1}(F_2)$ replace the code $(s; j, k)$ by $(s; j, n)$ and delete code $(\cdot; k, n)$ and $F_2(t)$.

Box 9 A test is made as to whether the current activity $(s; j, k)$ and associated activities $(\cdot; j, m)$, $(\cdot; m, k)$, $(\cdot; m, n)$ and $(\cdot; k, n)$ can be identified with the activities $(1; 0, 2)$, $(2; 0, 1)$, $(3; 1, 2)$, $(4; 1, 3)$ and $(5; 2, 3)$ of the Wheatstone bridge of Figure 3.

Box 10 The five c.d.f. functions involved on the Wheatstone bridge operation are combined in accordance with equation (3). The resulting $F(t)$ replaces $F_1(t)$, the code $(s; j, n)$ replaces $(s; j, k)$ and all other codes and c.d.f. are deleted.

The proof that the logic of the flow diagram in Figure 7 does indeed result in the computation of the c.d.f. of the critical time for any multiple-crossed network is given in the Appendix.

6. MONTE CARLO SOLUTIONS FOR THE MORE COMPLEX NETWORKS. As is well known and as was mentioned in section 1 the currently used PERT algorithm determines that path in the network for which the total of average completion times is a maximum. Now imagine that we apply the same algorithm to a random sample of completion times, each drawn from the distribution relevant to its activity. The 'critical time' so computed will be a single random variable from the distribution of critical times defined in section 1 and discussed in section 5. A large number of repetitions of this computation will therefore yield a Monte Carlo solution of the distribution of critical times. Such a solution will therefore be available for any network (and not just for multiple crossed networks).

Suppose now we are faced with a complex network (not necessarily multiple crossed). If we apply the algorithm of section 5 to such a network we would in general reduce the number activity - codes by the operations 'π', 'Conv' and 'Bridge'. However, if the network is not multiple crossed we shall not be able to reduce the network to a single activity. As soon as we find therefore that no reduction of codes has occurred on too consecutive cycles we would output the reduced network activities and associated c.d.f.'s so that it can be solved by Monte Carlo as indicated above. The operational calculus of section 5 will considerably reduce the complexity and extent of the network so that the subsequent Monte Carlo calculations are much simplified.

An IBM 709 computer program performing the above Monte Carlo computations of the distribution of critical times was prepared by L. L. McGowan (1964), in his M. Sc. thesis at the Institute of Statistics at Texas A&M University.

7. SENSITIVITY ANALYSIS AND GUIDE TO MANAGEMENT. The previous sections have been concerned primarily with the establishment of the mathematical, statistical, and logical aspects of determining the distribution of completion times for a project. The methods developed have further applications in analysing the effects of making specified changes in the original network and thereby providing guides for management actions. Basically, the analyses most readily recognized in this area are concerned with (1) assessing the impact of modifying the distribution of specified activities (e.g., a change in their average completion times); (2) assessing the impact of modifying blocks of

activities; (3) comparing two or more networks to establish the organization of the project for minimum time, minimum cost, or some other optimum; and (4) assessing progress or remaining time for the completion of the project.

All of the above assessments are permissible under the method developed in this paper. In fact once the logic is established on a computer, all four assessments are possible with the same computer programs. It is only necessary to vary the input and certain problem parameters according to the assessment required.

It should be pointed out that the assessments gained via this logic will be more comprehensive than a similar PERT assessment. With the present logic the impact on the c.d.f. of project completion times will be observable. This means that our sensitivity analysis provides estimates of the impact of production schedule changes on the expected completion time but also of the impact on its variance, percentiles, confidence intervals and other statistical parameters.

8. SPECIAL CASES OF BIAS DEMONSTRATION. As noted earlier bias enters the solution of a network problem due to inadequate treatment of the statistical considerations and approximate logic. In order to demonstrate this bias a few examples will be worthwhile for illustrative purposes. The following examples will also demonstrate the dependence of the solution on the distribution form and network composition.

EXAMPLE 1. Consider the case where k activities are in parallel as is illustrated in Figure 5. Assume further that each t_i is a random variable with exponential c.d.f.

$$F_i(t_i) = 1 - e^{-\lambda_i t_i}, \; i = 1, 2, \dots, k, \quad t \geq 0 .$$

The c.d.f. of the maximum time t is then given by

$$F(t) = \prod_{i=1}^{k} (1 - e^{-\lambda_i t}) \qquad (5)$$

If $\lambda_i = \lambda$ for all i, the mean of $F(t)$ is given by

(6) $$\mu = \frac{1}{\lambda} \sum_{i=1}^{k} \frac{1}{i} \ .$$

Clearly, since all $\lambda_i = \lambda$ and hence all $\mu_i = \mu^* = 1/\lambda$, the conventional PERT solution under this condition is $\mu^* = 1/\lambda$. The bias is then given by

(7) $$\mu - \mu^* = \mu^* \sum_{i=2}^{k} \frac{1}{i} \ .$$

Thus if there are only k=4 activities in parallel the bias will be $\mu - \mu^* = \mu^*(\frac{13}{12})$ or more than 100 percent of the PERT solution, whilst with k=8 activities in parallel the bias is 1.718 or 172%. It should of course be remembered that the above bias applies to the particular network in Figure 5 which, in general would only constitute a small section of the large network. Therefore, the % bias in the PERT - computed expected completion times will not, in general, be as large as the above example would indicate. However, PERT will <u>always</u> make underestimates of the critical time intervals (see e.g., Fulkerson (1962), p. 808) so that the biases from individual network sections will cumulate.

EXAMPLE 2: Consider the same network as above but with the density functions given by $f(t_i) = \frac{1}{c}$; $0 \le t_i \le c$.

In this case

(8) $$F(t) = (t/c)^k , \ 0 \le t \le c \ .$$

The mean value of F(t) is then

(9) $$\mu = \frac{kc}{k+1} \ .$$

The PERT solution would be the mean value of t_i which is $\mu^* = \frac{c}{2}$. The bias is found to be

$$\mu - \mu^* = \frac{k-1}{k+1}\,\mu^* \tag{10}$$

In this example the bias is at least bounded in that it cannot exceed 100% of the PERT solution. It does increase very rapidly however, with the number of activities in parallel. If k=4 as in the first example the bias is 60% of μ^*, when k=8 it is 78%.

EXAMPLE 3. To illustrate the dependence of the solution upon the form of the densities involved consider the following network.

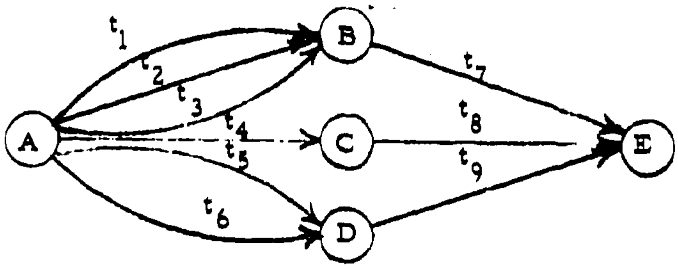


FIGURE 8
SHIFT OF CRITICAL PATH WITH FORM OF DISTRIBUTION

In this case suppose that the activities represented by the t_i have expected times as follows: -

Activity	Expected Time
t_1, t_2, t_3	9
t_4	11
t_5, t_6	10
t_7	3
t_8	6
t_9	4

If conventional PERT is applied, path ACE will be critical with a sum of expected times of 17 units. On the other hand, if the densities of the t_i are exponential and the operational logic of this paper is applied the expected time for ABE is 19 1/2 units, for ACE 17 units, and ADE 19 units, thus making ABE critical. This distribution dependence is further emphasized if the t_i are rectangularly distributed. In such a case the expected time for ABE is 16 1/2 units, and for ADE 17 1/3 units, thus making ADE critical.

The above examples, though somewhat elementary and academic, demonstrate the consequences of inadequate statistical treatment and approximate logic. The impact can be even more pronounced and the consequences more significant in a realistically large program plan.

9. RELATION TO THE EXISTING LITERATURE ON PERT. Most of the published work on PERT is concerned with computations based on the mean values of the completion times and deliberately ignores the bias discussed in this paper. There are undoubtedly situations when this bias is not serious notably in networks when

(a) There is a low degree of parallelism in the activities of the network and most operations are sequential and/or

(b) When some activities are carried out in parallel but one of them has a considerably longer expected completion time than the others parallel to it.

It will be agreed that the above conditions are not usually satisfied. In view of the very extensive, detailed and costly computations involved in the currently practiced PERT analysis it is surprising that so little attention has been paid to the bias affecting them.

We believe that whilst the possibility of a statistical approach (such as is here presented) has sometimes been considered (see e.g., Department of the Navy (1958), Appendix A, and Fulkerson, D. R. (1962)) it has apparently been regarded as leading to unsolvable or unmanageable mathematics. Indeed, Fulkerson (1962) who fully recognises the existence of the bias (see page 308) and offers an interesting approximate method to correct it, states (page 309): - "Since a typical PERT network may involve hundreds and thousands of arcs, the precise calculation of expected critical path lengths would, of course, be out of the question." Now it must of course be remembered that the method of numerical analysis here offered gives the solution only for the special case of multiple-crossed networks as here defined. We do not claim that the networks encountered in practice will usually belong to this category. However, if the algorithm described in section 5 is applied to a general network it will reduce it considerably so that the distribution of the critical time for the reduced network can be obtained by the Monte Carlo procedure described in section 6. Moreover, we could enlarge the scope of the numerical method of section 5 by adding (to the Wheatstone bridge operation for the network in Figure 3) similar basic crossed networks (such as that of Figure 4) and incorporate a calculation of the critical time (similar to that given by equation (3)) for such configurations. The feasibility and economy of such additions is under investigation.

Since we only give a hand full of references in spite of the vast literature on the subject, we should perhaps include the extensive Bibliography (Bolling Air Force Base (1963)) in our list.

REFERENCES

Bolling Air Force Base (1963). PERT Orientation and Training Center. Bibliography, PERT and Other Management Systems and Techniques.

Fulkerson, D. R. (1962). "Expected Critical Path Lengths in PERT Networks." Rand Corporation, RM 2075, P.R., Santa Monica, California.

McGowan, L. L. (1964). Monte Carlo Techniques Applied to PERT Networks. M.Sc. Thesis in Statistics, Texas A&M University.

Navy Special Projects Office (1958). PERT Summary Report, Phase I. (Gov't. Printing Office. Catalog No. D217.2, p. 94/958.

TEQUILAP: TEN QUANTITATIVE ILLUSIONS OF ADMINISTRATIVE PRACTICE*

Clifford J. Maloney

In 1949 I had occasion to install a punched card tabulator for the purpose of machine calculation of analyses of variance arising in a research, development, and testing program. At that time no need was felt to impose extraordinary restrictions on the procurement and utilization of such equipment. About the time of the beginning of the Korean War, however, higher authority, under the impression apparently that punched cards were employed only in Comptroller functions and that all such machinery was rented, instituted a requirement for monthly reports of per cent utilization, with a somewhat informal understanding that good management would secure a level of utilization of each piece of equipment at least as high as 50% and that 80% would be much more appropriate. Having encountered what has since come to be called queuing theory in Thornton Fry's text on probability theory many years earlier, I had a summer worker in 1955 make an application of these considerations to the congestion delays which would result from any given level of per cent utilization. These are shown in Figure 1. This study has appeared in a paper given at the Second Statistical Engineering Symposium at Edgewood Arsenal in April of 1956. However, my efforts and those of others--a few of which have come to my attention--to point out the costs as well as the benefits as per cent utilization increases had an absolutely zero effect on "administrative practice." Two major conclusions from this experience and many others, before and since, were, however, made clear to me. The first conclusion is that decision making is an emotional, not a rational, operation. People often bolster their decisions by arguments--some of them rational--but seldom reverse the process. I am sorry to say that so far as I can see this holds as much for logicians and mathematicians as for anyone else. This is of course what is meant by that well known saying: "I've already made up my mind; don't bother me with the facts." The second conclusion deals with the arguments by which it is customary to rationalize emotional decisions. Even where the "reasoning" can be accepted as not totally irrelevant, it will be based, not invariably but very often, on unwarranted but unquestioned assumptions. My example

*The views expressed herein are those of the author and are not to be ascribed to any other agency or individual.

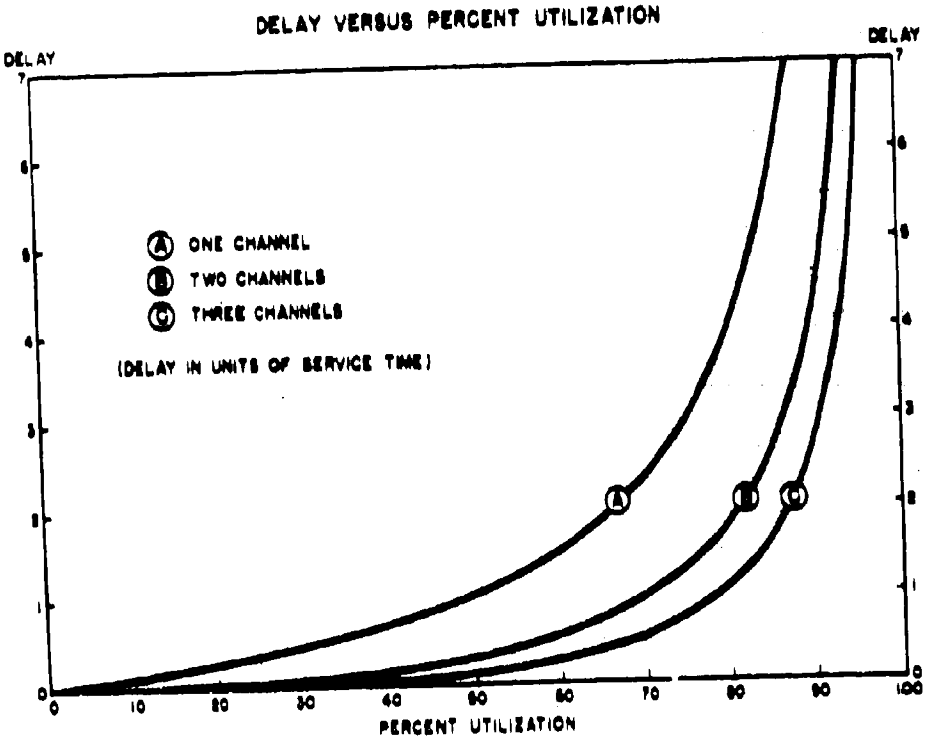

DELAY VERSUS PERCENT UTILIZATION
DELAY
DELAY
Ⓐ ONE CHANNEL
Ⓑ TWO CHANNELS
Ⓒ THREE CHANNELS
(DELAY IN UNITS OF SERVICE TIME)
PERCENT UTILIZATION

Figure 1

of per cent utilization as a measure of punched card installation efficiency illustrates this. One assumption was that the equipment was rented and not owned. Another, that efficiency was not merely a function of utilization but actually a monotone function of it.

I cannot claim originality in this insight. In a commencement address at Yale University, June 11, 1962, the President of the United States said: "For the great enemy of the truth is very often not the lie--deliberate contrived, and dishonest, but the myth--presistent, persuasive, and unrealistic. Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. We enjoy the comfort of opinion without the discomfort of thought . . . . Mythology distracts us everywhere--in government as in business, in politics as in economics, in foreign affairs as in domestic policy." The former President's indictment is much stronger and more inclusive than mine.

Allyn Kimball* has defined "errors of the third kind" as giving correct answers to the wrong questions. The assumptions of the question become postulates of the answer. He observes that a first step in finding useful answers is to query the question. The originator of the cognate insight that most of what in common life passes for argument consists of more or less accurate deduction from wrong premises is lost in the mists of time. But perhaps there is room for me to assert some small claim to originality in the recognition that many of these false premises spring from an inability or an unwillingness to think in quantitative terms, to see the clarifying role of an appreciation of their quantitative nature, and to see that some at least of these false deductions cannot be resolved by logic alone; as their very nature is inherently quantitative.

The application of mathematical principles in administrative practice which I illustrated in my first example, specifically an application of probability, is relatively sophisticated. Further, the problem which gave rise to it, while important, was rather limited in scope. Most of what administrators do day in and day out consists of more homely if more important actions, though it is true that queuing theory has many applications there

---

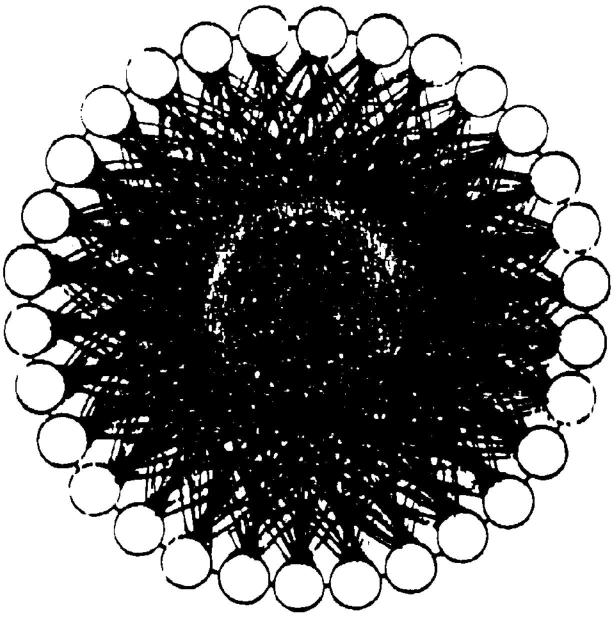
* Allyn Kimball, "Errors of the Third Kind in Statistical Consulting," JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, June 1957, p. 133.

also. Dr. Edward F. R. Hearle* in an article "How Useful Are 'Scientific' Tools of Management?" enumerates them as: "linear and dynamic programming, queuing theory, game theory, simulation, and monte carlo, to name a few." The general tenor of his appreciation of these tools and, hence, of quantitative thinking in management is believed expressed in his sentence: "Furthermore these tools do not deal with some of the more exciting parts of the total management process . . . . ." I take exactly the opposite view to the one that I believe Dr. Hearle is espousing; that quantitative thinking is not (as he asserts) limited to formal manipulation of numerical quantities but is very useful where the mere recognition of a variation from instance to instance of a given type is involved.**

The serviceability of adhering rigidly to the "channels" of an orginization chart can be judged by indicating the lines of contact that exist in the absence of "organization." (Figure 2). The organization chart of Figure 3 was selected, not to suggest that administration goes around in circles, but because the chart is round and, hence, emphasizes the contrast with Figure 2. The organization replaces an unorganized conglomerate. Now, when the production department gets ready to fill an order, they send the goods and the invoice to the President's office, and his secretary passes them on to the shipping and accounting departments. At least no organization chart of which I am aware gives any guidance to the contrary. This may be the reason that organization charts and the "authority lattices" behind them have such a low reputation outside of organization and management departments. There is a great deal of discussion and some useful research which distinguishes between "formal" and "informal" organizational relationships, but a recognition that a great gulf so frequently exists

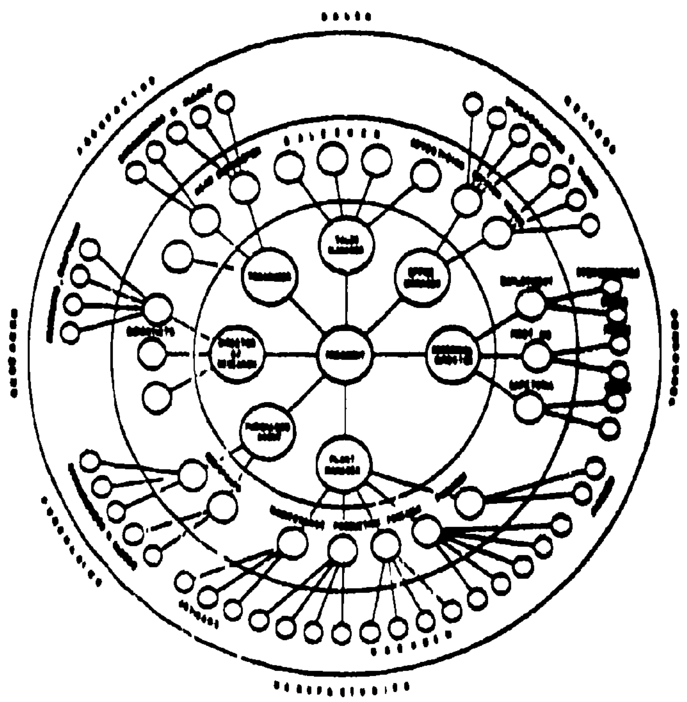
*Dr. Edward F. R. Hearle, "How Useful Are 'Scientific' Tools of Management?" PUBLIC ADMINISTRATION REVIEW, Autumn 1961, pp. 206-209.

**In making his statement Dr. Hearle had the type of mathematical tools which he had enumerated, and others like them, in mind. So my challenge to him relates to the inference which can fairly be drawn from his remarks rather than to any direct statement which he makes.



DISORGANIZATION CHART

Figure 2



Organization Chart *

Figure 3

*From COOPERATION AND CONFLICT IN INDUSTRY by F. Alexander Magoun (Harper & Brothers, 1960).

between these structures is itself a witness of an inadequate formal structure in the organization--as is the fact that the disparity is so widespread a testimony to the unrealistic (and unhelpful) nature of current theories of organization.

The comparatively sterile status of organisation theory may be due to the inherent difficulty of the subject, the low ability of those who work in the field, or to the absence of one or a few essential concepts not yet sufficiently clearly delineated. The author is encouraged to propose that the latter may be an essential feature, and that the principles to be discussed may be included among the missing essential elements. Few of these points are entirely original, yet essentially none are clearly and widely appreciated.

Reflection of many years, exponentially enhanced in intensity in recent months, has led me to subsume the most significant examples of quantitative illusions in administrative practice which have come to my attention under 10 headings, partly because I have 10 fingers, and partly because this produced the acronymic title. This paper will consist of a listing of these 10 principles with a few exemplifications--not to convince anyone of the truth of my position on the examples; it would be impossible to discuss more than one or two in the little space allotted to me--but to demonstrate the clarification let into many administrative forms of action which otherwise must remain, as they were formerly to me and must still be to you, an enigmatic mystery. If the reader feels that my examples are inadequate or wrong, he is invited to refer to others from his own experience. Only if he feels that none or few can be found does he have a quarrel with my principles per se.

My first "quantitative illusion of administrative practice," (Figure 4), is entitled "Peas in a Pod," to emphasise that (a) the assumption that is generally made is that "if the name's the same, the thing's the same"--"as alike as peas in a pod," and (b) that the assumption is wrong. I see this as the exact opposite of Professor Hearle's (implied) position cited earlier. We can gain greatly by merely recognising this fallacy, without in any way being able to quantify it. This is at once the most important and the most fundamental of my 10 illusions. The fallacy lies in the denial of the reality and significance if quantitation in situations where it is real and important. Sometimes it takes the slightly more subtle form of acknowledging: yes, the several members of any one class do

PEAS IN A POD

Figure 4

differ and perhaps differ widely, but it is administratively impossible to allow for every possible variation; therefore, we will allow for none. The same punched card computing installation provided a glaring example of this "reasoning." We needed a card punch operator. The position analyst referred to a job standard which explained that card punch operators work from edited repetitions item records which are punched mechanically with no exercise of individual intelligence or ingenuity. The position analyst had no need to look at the facts when he had a book to tell him that (a) this position was identical with all other positions called by the same title, and (b) one such position filled the book specifications. Another exemplification of the illusion which receives a great deal of public attention is lowest bid procurement. Since all items (even those not yet invented) and all services can be exactly described in the invitation to bid, all are equivalent; and hence the purchase price is the one remaining variable. Of course, there is a contrary cliché: "you get what you pay for." There is no requirement in "administrative practice" that the system of clichés be consistent.*

My second illusion (Figure 5) attempts to deal with the assumption that all the good qualities reside in one product or one course of action, and, by inescapable logic, all the bad lie in any alternative--though if there are several, these bad qualities may be distributed among them. I am sorry to say that I was supplied with a perfect example of this

*It is a truly remarkable thing that philosophers, since the time of Plato, have been concerned with the problem of "nominalism" versus "realism," which, however, important theoretically, seems not to constitute a stumbling block in day to day relationships; whereas this first illusion is at once the most pervasive and most pernicious logical fallacy entering not just into almost every discussion between friend and foe, between advocate and adversary, but between even so intimately related and favorably disposed groups as members of one family. It was the essence of the "hyphenated-American" dispute, the merits and the abuses of political party labels and party loyalties, methods versus subject matter in education and of occupational jurisdictional disputes, whether within one organization or between competing parties or groups.

"Maybe she'd catch up if you didn't put so much stress on the importance of accuracy over speed."

PROFIT AND LOSS*

Figure 5

*Used by special permission of KING FEATURES SYNDICATE.

illusion in the 16 October 1964 issue of SCIENCE. Two of this country's most illustrious scientists explained their choice of candidate for President. I have examined these statements with some care. Neither protagonist could find a fault worth mentioning in his own choice nor a good quality in the latter's opponent. This action constitutes a conformance to the practice in disputation.* But it is the fact that the practice prevails in the day to day administrative process that concerns us. Why does it? The complete explanation must lie in psychology. A plausible treatment of just this phenomenon has been contributed by Professor Leon Festinger of Stanford University.** Of course, Professor Festinger is not responsible for my understanding or use of his theory. In essence, the mind demands harmony. Yet all real things and all real courses of action involve advantages and disadvantages. The only achievable harmony is a quantitative one--a balancing of opposing forces. But this type of harmony is uncongenial to many minds.

That, if one embraces one course of action or one belief, he must impute all virture to the chosen, and all evil to the rejected, was long ago recognized as a fundamental error by Georg Wilhelm Hegel, last of the global philosophers, who saw progress of social organization in the reconciliation of the thesis and the antithesis into a synthesis that removed the conflict by absorption of the thesis and antithesis as elements in a higher concept. Why Hegel never "caught on" in administrative practice I cannot say. But it is possible that, since his view was essentially qualitative and not quantitative, hence didn't in fact apply in many instances, his solution tended to be neglected even when perfectly applicable. An entirely analogous situation is known to have delayed acceptance of the contagion theory of disease for centuries.

---

*One of the authors was good enough to acknowledge receipt of an early draft of this paper with the statement that the allocation between the two discussions was a deliberate attempt to conform to the anticipated expectations of the readership of SCIENCE. I do not have the boldness to point out that insofar as the anticipation is correct, my strictures apply then to the readership if not the disputants.

**Leon Festinger, "A Theory of Cognitive Dissonance," (Evanston, Illinois, Row, Peterson, 1957) pp. 260-285.

Those who escape the first illusion and recognize that quantitative variation pervades most, if not all, of life are candidates for the notion that, if a little is good, more is better. (Figure 6). The "spartan" philosopher would put that from the standpoint that, if moderation is good, abstinence is ideal. I have seen this illusion active in the question of where to put the statistician in an organization. The same holds for engineers, stenographers, air support, artillery, computers and machine tools. If the drawbacks of the widest possible organizational scattering are recognized, then complete centralization appears inescapable--and vice versa to most administrators, organization specialists, operations analysts and other people who move productive workers around. The most extreme proposal that has come to my attention is to centralize all computers in government. The one facet of all these matters that is of concern here is the human tendency, once embarked on a path, to assume that that path leads upwards (or downwards) indefinitely. The visible existence of "side effects" in therapeutics has compelled an avoidance of this illusion with full virulence in medical practice--but, as recent history shows, the tendency certainly exists.

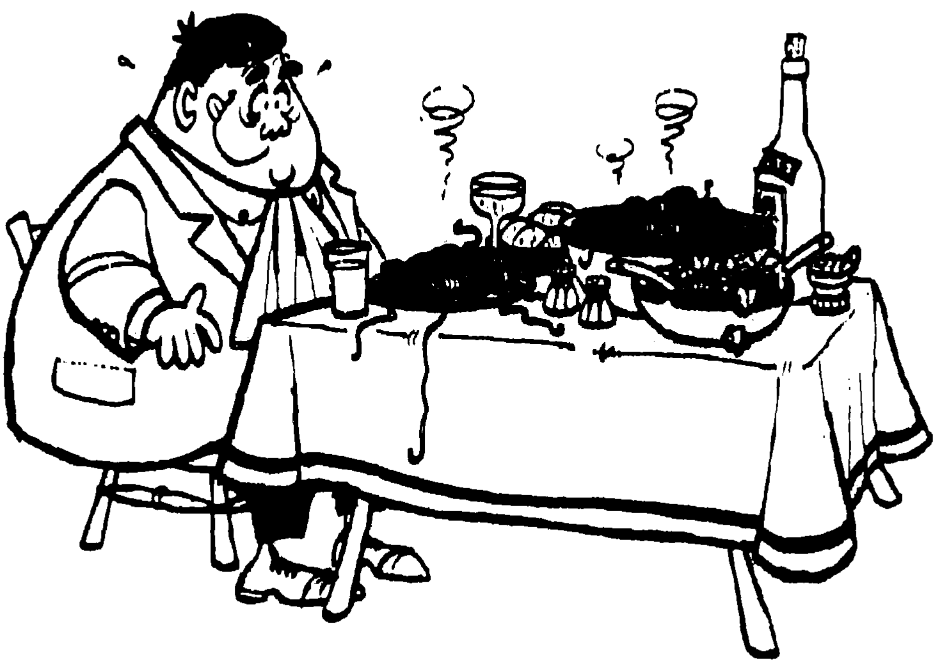
The recognition of the existence--though I think not the nature--of this tendency has led those who prefer the status quo to take refuge in warnings against "the foot in the door," "the opening wedge," "the breach in the dike," "the camel's nose under the tent," with the consequence that anyone who fears extremism acts like and is characterized as an obstructionist. A recognition that the action or state rebelled against would lose its terrors (so that its advantages could be secured) if illustion three could be eliminated might remove much acrimonious social debate.

Forty years ago there was a principle in psychology, that I have never heard mentioned since, which involves a modified form of this fallacy and provides a partial explanation for its existence. I can best exhibit its nature by recalling to mind the once popular medical treatment of bleeding. General Washington was bled in his last illness and the practice continued, though with declining popularity, to the Civil War, and even lingered till World War I. Suppose a patient is "treated" by bleeding but succumbs. Three conclusions are possible. The bleeding either was (1) deleturious, (2) indifferent,

Those who escape the first illusion and recognize that quantitative variation pervades most, if not all, of life are candidates for the notion that, if a little is good, more is better. (Figure 6). The "spartan" philosopher would put that from the standpoint that, if moderation is good, abstinence is ideal. I have seen this illusion active in the question of where to put the statistician in an organization. The same holds for engineers, stenographers, air support, artillery, computers and machine tools. If the drawbacks of the widest possible organizational scattering are recognized, then complete centralization appears inescapable--and vice versa to most administrators, organization specialists, operations analysts and other people who move productive workers around. The most extreme proposal that has come to my attention is to centralize all computers in government. The one facet of all these matters that is of concern here is the human tendency, once embarked on a path, to assume that that path leads upwards (or downwards) indefinitely. The visible existence of "side effects" in therapeutics has compelled an avoidance of this illusion with full virulence in medical practice--but, as recent history shows, the tendency certainly exists.

The recognition of the existence--though I think not the nature--of this tendency has led those who prefer the status quo to take refuge in warnings against "the foot in the door," "the opening wedge," "the breach in the dike," "the camel's nose under the tent," with the consequence that anyone who fears extremism acts like and is characterized as an obstructionist. A recognition that the action or state rebelled against would lose its terrors (so that its advantages could be secured) if illusion three could be eliminated might remove much acrimonious social debate.

Forty years ago there was a principle in psychology, that I have never heard mentioned since, which involves a modified form of this fallacy and provides a partial explanation for its existence. I can best exhibit its nature by recalling to mind the once popular medical treatment of bleeding. General Washington was bled in his last illness and the practice continued, though with declining popularity, to the Civil War, and even lingered till World War I. Suppose a patient is "treated" by bleeding but succumbs. Three conclusions are possible. The bleeding either was (1) deleterious, (2) indifferent,



SO MUCH THE BETTER

Figure 6

or (3) helped but was inadequate. In the third case, the remedy is even more heroic bleeding. This latter conclusion is associated most strongly with the name of Dr. Benjamin Rush, Professor in the first medical school in America.

In the psychological context this process--the persistence in a wrong course of action under the misapprehension, entirely sincere, that, while the strength was weak, the sense was right--was called the beta hypothesis. It seems most curious that such a fecund insight should drop from view.

But that it is more important to run in the right direction than to gain great yardage was redemonstrated in a football game Sunday, 25 October 1964, by Minnesota end Jim Marshall. A gyroscope persists on the course of its setting despite contrary forces. But does that make the setting right? Methods for determining the "direction of choice" have long been known in the science of statics and have, more recently, been investigated in economics. The glaring fact that scholars cannot agree on the direction of the consequences of an economic action is a greater blow to the status of the science of economics than the uncertainty of the magnitude of the effect in those few cases where there is agreement on the direction.

Innovators since the beginning of time have regarded all who saw merit in the old as obstructionists. That they may merely be victims of the beta virus has, so far as I know, never been entertained. Galileo's experience is the most famous, if not the most meritorious, example, but the theme is commonplace. It is a reliable story to tell in the movies, on TV, and in the pulp press. In real life it is as common in criminal prosecution, apparently, as in scientific innovation.

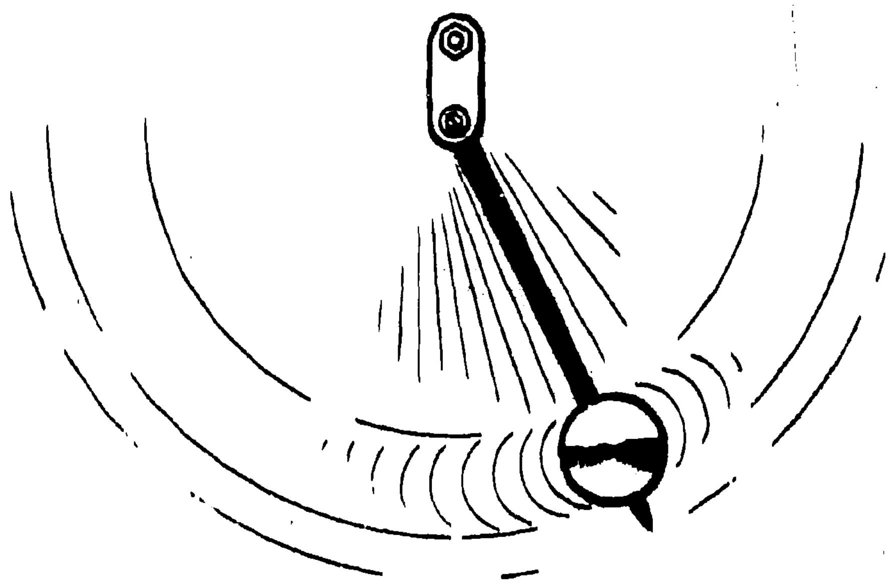
In the common view, stubborn insistence on one view and a refusal to even hear the evidence for another is a sign of malevolence on the part of the defenders of the status quo. In the psychological beta hypothesis it was a particular form of mental illness, possibly mild in character and little disabling. In a seemingly neglected paper in SCIENCE, the philosopher, Michael Polanyi, citing his own severe injury from exactly this process nevertheless defends it as essential to progress in science. He says: "there must be at all times a

predominantly accepted scientific view of the nature of things in the light of which research is jointly conducted by members of the community of scientists. A strong presumption that any evidence which contradicts this view is invalid must prevail. Such evidence has to be disregarded even if it cannot be accounted for, in the hope that it will eventually turn out to be false or irrelevant."

This thesis of Professor Polanyi seems to be receiving just the treatment which he argues it must. Despite his evident eminence, both as scientist and as philosopher, he appears to be suffering from an acute case of the second (Profit and Loss) illusion, complicated by the presence of the beta virus. He sees that science, functioning as it does, advances. He does not see how it could do so were it to purify itself.

It is clear that the seriousness of the third illusion is unmistakable, once its reality is granted, but when combined with the second illusion (that one of alternative courses of action, degrees of centralization, size of computer, has all the virtures and none of the faults--or that one must at least act that way) explains, I believe, the peculiar property of progress in administration. (Figure 7) At one extreme, (form of organization, method of scheduling work, approval channels, work procedure, etc.) the adherents of the other will have ample proof of the inadequacy of the chosen solution--and they will be right. Fears of the "opening wedge" effect will, however, maintain the existing status quo until proponents of the opposite extreme gain sufficient ascendency to overcome these fears, when a shift towards a balanced solution will set in. Now, just when the optimum position is reached, the third illusion will add momentum to the swing until the opposite extreme is attained.

In consequence of the interaction of these two illusions, "Profit and Loss" and "So Much the Better," progress in administrative practice--and many other facets of human behavior--consists in discarding yesterday's procedure and adopting that of the day before, with the assumption and, indeed, the claim that the new is novel. There is no risk; for few will have survived from the earlier period, and they can be suppressed. The number who, not personally surviving, will have read history will be of measure zero.



ADMINISTRATIVE PROGRESS

Figure 7

Professor Jay W. Forrester of M.I.T. argues* that this result is in fact a necessary and not accidential characteristic of administrative progress under the existing circumstances. He writes: "In the past, management methods have been learned primarily through personal experience. The developing manager rotates through numerous assignments. Management schools repeat the folklore and the experiences of practicing managers. This experience is used as a basis for generalizing, so that past experiences can become a basis for anticipating the nature of new situations."

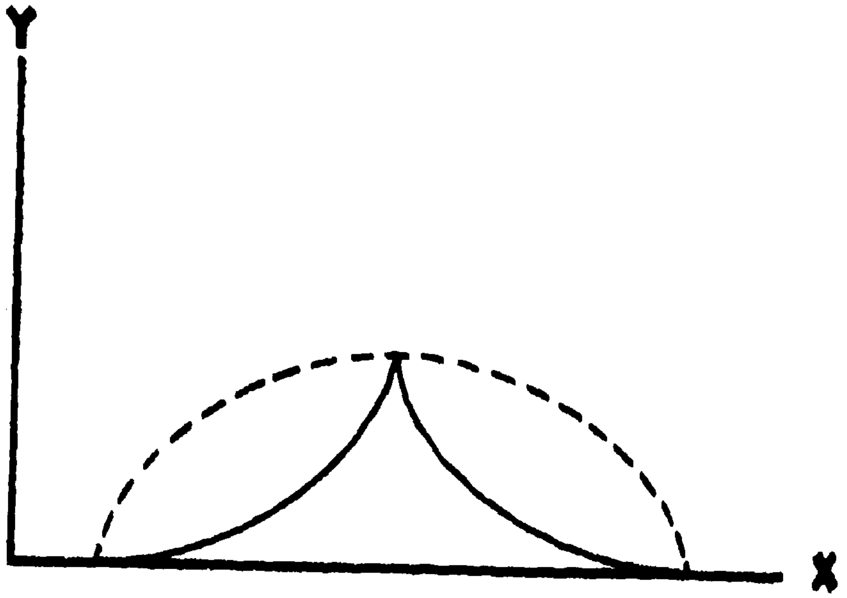
". . . we have here at the M.I.T. School of Industrial Management been developing an approach to management policy design which we call 'industrial dynamics.' It is intended to be a new way to understand how corporate structure and policy produce the different characteristics which one sees in business enterprises. . . . Most managers are surprises to learn that those practices which they know they are following are sufficient, when assembled in a system model, to cause the major difficulties which they have been experiencing."

Professor Forrester is too polite to do so, but I cannot resist observing that the jibe: "He has never met a payroll" is a pistol pointed backward.

Figure 8 shows another illusion widely prevalent but not recognised as quantitative. It is closely akin to illusion one, but differs in that here the underlying phenomena is recognised as quantitative, e.g., different grade levels for different duties, adjustments in the general level of compensation, fringe benefits, quality of tools or equipment, work space, and so on, but it is assumed that all maxima are cusps; that if some precise (even if unknown) value is optimum, then that even slight departures result in great waste of resources. My own chief experience in this field had to do with the erection of a building to house a computer and associated staff. We were asked to build for the future, but to justify in great detail just how every square foot would be assigned.

---

*Jay W. Forrester, "Dynamics of Corporate Growth." Paper delivered at a conference on "Management Strategy for Corporate Growth in New England," held at M.I.T. November 12, 1963. This brilliant paper deserves reading in its entirety for its positive approach to rescuing administrative practice from the grip of injurious if plausible "folklore."



THAT'S ENUF, CUSP IT

Figure 8

The interests of the government would be served if we were adequately housed, but, though the building would have a life expectancy measured in decades, it seemed to be assumed that any attention to contingencies beyond the needs of the instant, (e.g., an anticipation of the tortuous delays in the path to actuality) whether of our unit or of other units on the Post, could only be "wasteful." Opposition to grade escalation; to better than "necessary" research facilities, military aircraft, weapons, uniform belt buckles, or missile boosters fall in this category. If I say that I think this line of "thought" led to the Russian scoop in space, the importance of this illusion, if not its actuality, will be obvious.

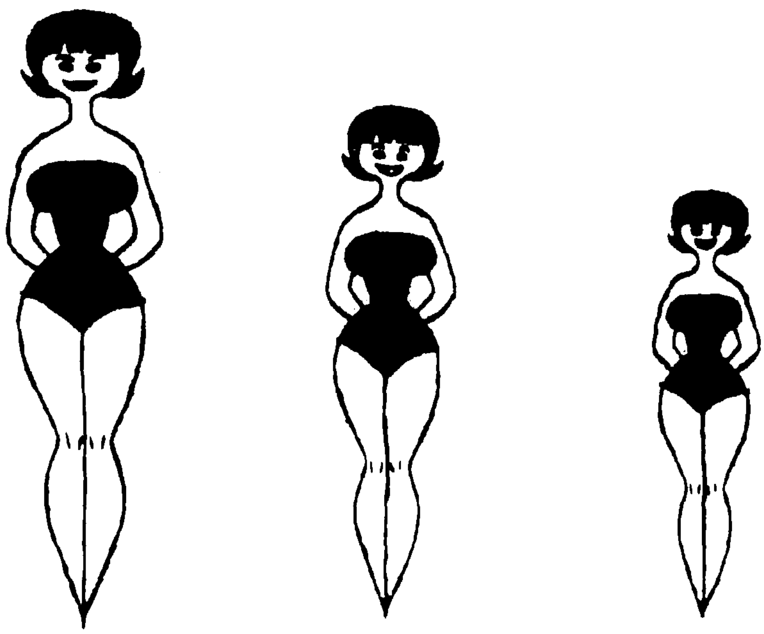
In his treatment of "The Economics of American Medicine" Seymour E. Harris,* it seems to me, suffers severely from this illusion when he refers to the "waste" of better than "necessary" medical attention or hospitalization.

Illusion five (Figure 9) is more conventional. At first blush one would assume that the factor which varies between the three figures is one of height only, despite the fact that we all know that the human figure is a three-dimensional object; hips, waist, and of course, height. Perhaps others. I include this illusion only to pay homage to a factor well recognized in all circles, mathematical and anti-mathematical, the multi-dimensionality of real life problems. The reader is asked to stretch a point in the interest of economy of illustration and view the figures as also suggesting the non-linear (curvilinear) character of most realistic situations. Again, the availability of the computer could push these difficulties back a step or two, were the nonquantitatively oriented administrator aware of the potentialities.

This illusion is unique in that it occurs with great frequency in two opposite forms. The tendency to "oversimplify a problem" is widely indulged--but widely condemned. Indeed the opposition to automation, to "thinking by computer," while reaching a crescendo in modern times, has had its Cassandras in all ages.

---

*Seymour E. Harris, "The Economics of American Medicine" (New York, Macmillan, 1964).



A ONE DIMENSIONAL VIEW

Figure 9

These anti-mathematicians assert that no amount of complicating the model can hope to mirror reality and that the administrator deals with reality. There is a complementary consequence. The administrator, like the observer, "never knows what he is talking about or whether what he concludes is true," and without recognizing it, he has conceded as much by the above argument. Philosophers of the scientific method have repeatedly pointed out the contrast between the rigor of conclusions derived from contrived but designed experiments and from passive observation of a complex, if real, world.

A particularly striking example of this phenomenon appeared in a recent column of Walter Lippmann, concerning the place in history of Herbert Hoover. Lippmann says: "we avoided such a crash [the Great Depression] after the Second World War because we had so well learned the lessions of the First World War." Clearly, we didn't learn the larger lesson of how to avoid war. And is our continuing prosperity a reward for a lesson well learned or a penalty for a cold war still in progress?

A widely publicized illustration of this phenomena has arisen in the controversy over smoking and disease. The evidence in the human is observational and not experimental. The same administrators who most adamantly cling to this objection to an unwanted conclusion, never hesitate to advance the sales force on the basis of their "proven record." That these two attitudes are inconsistent is my only point.

All soothsayer type forecasting of the future, including forecasts of election outcomes, depends for its acceptance on the fact that the public suffers from the misapprehension that one success is a guaranty of superhuman powers. Business chooses its tycoons, and government its "bright young men" by this technique. I hasten to add that it guarantees the wrong choice no better than it does the right. It may be called the one step (two step, k-step) beta decision rule.

The complexity that defeats description denies certainty. If the administrator must act today, guided only or principally by his "intuitive grasp" of the "total situation" and not on the basis of previously enunciated rules thought to embrace the situation at hand, then he cannot carry over any lesson for tomorrow whether he is successful or a failure today.

That experience confirms error as often as truth is hardly a novel observation. This is in fact the beta hypothesis. One does indeed learn by experience, but at present only by an extravagant volume of repetition, which then forces an abstraction into the consciousness of the "man of action." Once verbalized, experience can confirm or refute the hypothesis. Again, the beta hypothesis asserts that experience will lead to the explicit formulation of wrong inferences quite as much as right. But just as bleeding as a medical treatment finally gave way to overwhelming contrary evidence, so must any belief in conflict with reality, if only there is sufficient experience since, ultimately, despite the influence of the forces described in the beta hypothesis, there will be an attrition of incorrect deductions and an enrichment of correct ones. There is a statistical technique to deal with such cases--Sequential Analysis--but its applicability in administrative contexts must so far have occurred to only one individual.

True, the practical administrator would never assemble the number of observations suggested by Sequential Analysis before making his decision--except in matters of trivial moment. He is thereby protected by (and confirmed in) his belief that his poor batting average belies the very quantitative reasoning he ignores!

Illusion six (Figure 10) is the most pervasive of all, for we each see the world only from our own point of view. In Adam Smith's day it was argued that the interplay of universal self interest would produce a general maximum of social well being through the action of an "unseen hand." It has since become clear that the hand is either unsteady or unfriendly. But the same illusion has taken refuge in the cry "let the experts decide." Curiously, it is in the military field where the dictum of Clemenceau: "war is much too important to leave to the generals" is most known and most accepted. Clemenceau should have realized that if he didn't say that road building is too important to leave to the engineers and management too important to leave to the managers that no successor would arise capable of that generalization. The experts should be left to their own devices--when they are concerned with matters of interest only to themselves. In particular, "research managers" are only effective when they adhere to Jeffersonian principles of government (not his actions). What happens when diverse activities impinge--each the field of a different group of "experts"--is the subject of illusion seven.

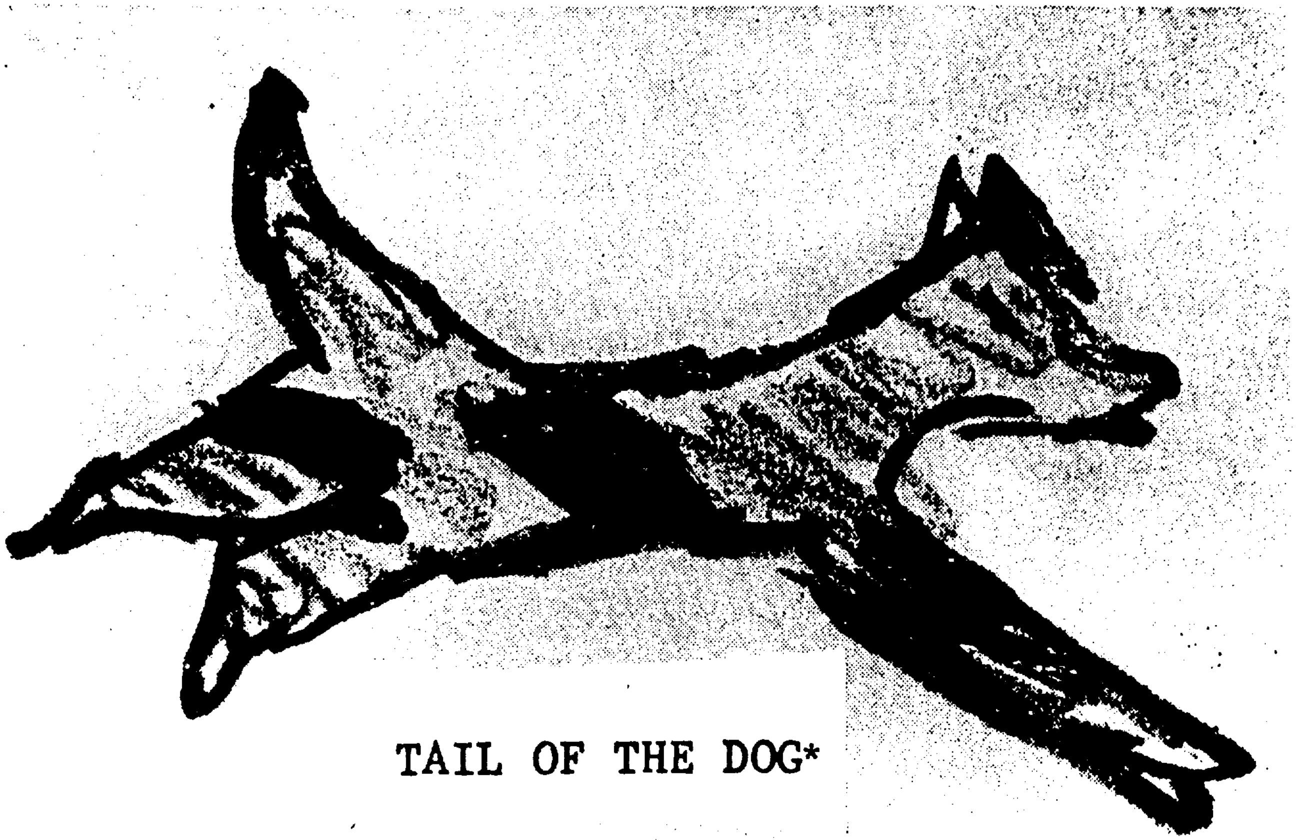


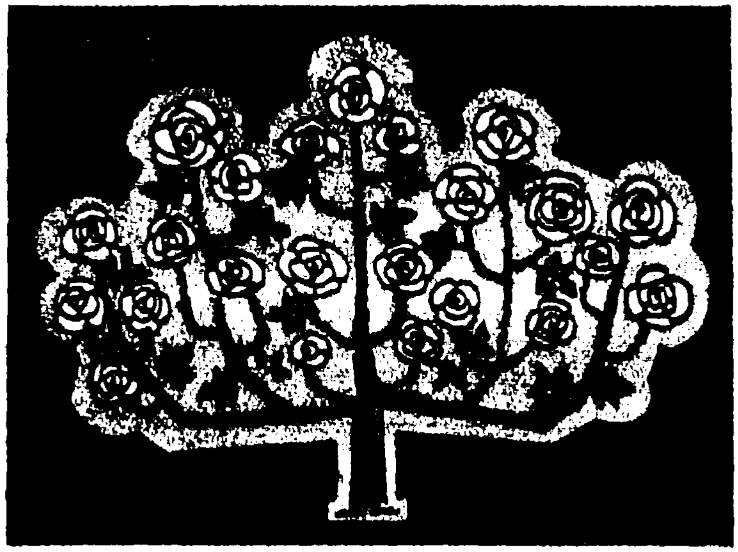
Figure 10

The "blight of the expert" takes two forms. In the previous paragraph I discussed the effect of letting the expert "do his own job" where it, however, impinges on others and/or on their capacity to do theirs. The second form is to extend the expertise from the field in which it is earned into other distinct fields where it is not. This illusion is related to but quite distinct from illusion one. In "Peas in a Pod," the assumption is that, for example, every engineer is interchangeable with every other engineer. This second form of the "Tail of the Dog" illusion is that any engineer (physician, military officer, educator, or bricklayer) and only an engineer (physician, military officer, educator or bricklayer) can perform any other function regardless of the skills involved in an engineering (medical, military, educational, or construction) organization. The example that caught my eye here is: * "I think all good statisticians agree--and I define a good statistician to be one who agrees--that statistics is not mathematics. On the other hand, it happens to be a peculiar subject of its own, which mathematicians when they do take the trouble, can teach much better than non-mathematicians."

Illusion seven (Figure 11) is intended to conjure up in the viewer's mind not just the genus Rosa but the whole field of classification. Classification is associated most securely in the popular mind with the fields of zoology and botany--but is somewhat less frequently recognized as a universal tool of science. Languages, rocks, and forms of tribal kinship relations are classified. The most famous classification in mathematics is that of Klein, but others continue to arise in every quarter. Second only to biology the public comes in contact with classification at the public library. Librarians, at least in America, will stress that a universal classification is not obtainable--but administrative practice has not yet learned this lesson. The form in which this illusion enters here is in the search for the holy grail of the perfect organization.

---

*John G. Kemeny in "New Directions in Mathematics," proceedings of conference arranged by John G. Kemeny and Robin Robinson. Edited by Robert W. Ritchie. (Englewood Cliffs, New Jersey, Prentice-Hall, 1963). I strongly suspect that Professor Kemeny made the above remarks with tongue in cheek and has succeeded in pulling my leg. Why else would he commit such a transparent logical non sequitur?



A Rose is A Rose is A Rose

Figure 11

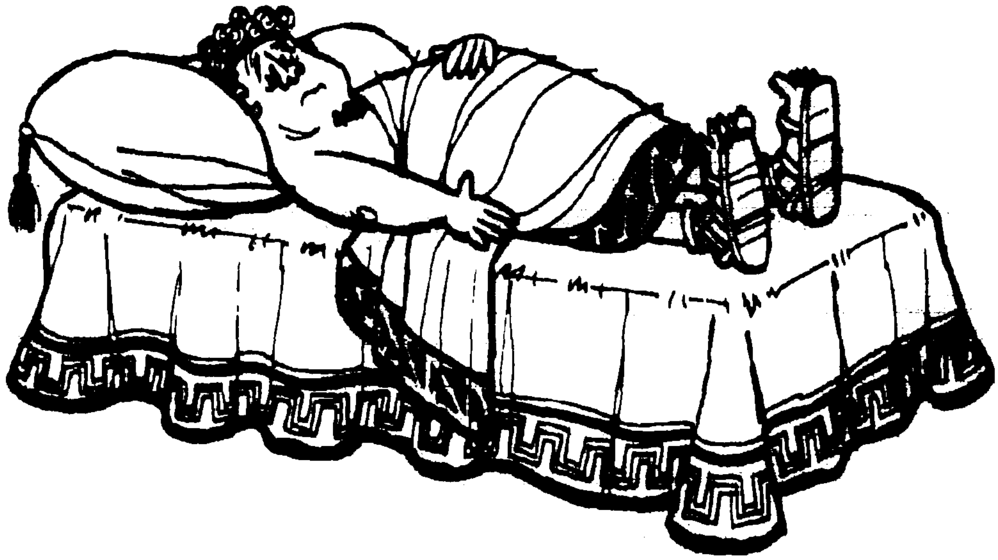
That a satisfactory organizational classification for the distribution of responsibilities has not yet been achieved is shown by the fact that we are forever reorganizing. What I wish to call attention to here is that this process is never interrupted to check on the existence theorem, nor to seek a fruitful setting for an examination of the problem. Organisation is the replacement of the disorganization chart by the organization chart. Bureaucracy is the conversion of the paths in the disorganization chart from paths of persuasion to paths of coercion. One testimonial to the fact and to the consequences was the establishment by the Army of the "Program Managers."*

In the Navy's Polaris Program, (the one instance success which convinced the Army) according to an article in the Civil Service Journal for July-September 1964, Admiral Raborn was authorized by letter to get "whatever people and whatever cooperation he required from any of the Navy's bureaus and offices." Yet "Admiral Burke admonished him that if the letter ever had to be used to force cooperation, the project would fail." Notice that Admiral Raborn went out of channels, he used the organization of the disorganization chart, he used these lines as lines of persuasion--not as lines of control. This is where organization began!

Procrustes' bed (Figure 12) was the ancients' way of putting the principle of what we now know as "the organization man." In America every man is an individual--"just don't step out of line." I wish to unite this principle with another not perhaps widely known but at least not due to me. It has been remarked that never in history (of course; seldom in history) have the professionals invented a radical departure destined to lose them their jobs.

In an article, "Revisionist Theory of Leadership" by Professor Warren G. Bennis in the Harvard Business Review for January-February 1961, Volume 39, pages 26 ff, the author on page 148 observes: "along these lines, Samuel Goldwyn was reputed to have said to his staff one day, 'I want you all to tell me what's wrong with our operation--even if it means losing your job!'" I thought in Hollywood telling the boss what he was doing wrong was a sure way to lose one's job. I'm sure that

*The beta virus strikes again! It is something of an achievement, surely, that bureaucracy is to be cured by increased bureaucracy.



PROCRUSTES' BED

Figure 12

Professor Bennis, like me, thinks this story is aprocrophyl--but it at least lists one impediment to people with an assigned mission ever making very radical changes in how its done.

The National Road starts at Braddock's Rock near the Lincoln Memorial and runs through Frederick, Maryland, and then on west. It has been said that the waggoners on the road didn't develop the canal system, the latter didn't develop the railroad. No railroader developed the automobile or the airplane--indeed, didn't even develop the diesel locomotive. But administrators have not yet drawn the obvious conclusion. If a department is charged with a certain responsiblity, then that department will never introduce radical changes. If anyone does, someone else will do so, But so soon as he does, he will have the organization manual thrown at him. Anything he may have done will be destroyed and the task will be transferred from one who cares and knows to one who fears and opposes. At least that was my experience, and of several of my former colleagues.

This consequence is particularly unfortunate, since if my claim that the "research manager is best who manages the least" is correct, then this is the environment which will yield the largest number of outstanding results. But the consequence of illusion seven (A Rose Is a Rose Is a Rose) will be that such a manager will get a low rating and be accused of permitting "excessive duplication." Even here I can claim no priority. Albert Hirschman and Charles Lindbloom in an article "Economic Development Research and Development Policy Making: Some Converging Views," in BEHAVIORAL SCIENCE, Volume 7, 1962, pp. 211-222, explicitly recognize the benefits of a little "play" in the tight constraints placed onresearch and development efforts.

Illusion seven (A Rose Is a Rose) and eight (Procrustes' Bed) together are at the bottom of much of the ferment over "management" of research. It seems so simple to divide up the field of action and portion out support, responsibility, and facilities to each. We may even take a leaf from the military field commander, and recognizing that liaison between adjacent commands is the weak point in a battle line, we can be tolerant of a certain amount of interpenetration at the peripheries. But innovation arises from the darnedest sources. If it arises in the civilian economy or at least from non-governmental sources, we need merely create an Office of Scientific Research and

Professor Bennis, like me, thinks this story is aprocrophyl--but it at least lists one impediment to people with an assigned mission ever making very radical changes in how its done.

The National Road starts at Braddock's Rock near the Lincoln Memorial and runs through Frederick, Maryland, and then on west. It has been said that the waggoners on the road didn't develop the canal system, the latter didn't develop the railroad. No railroader developed the automobile or the airplane--indeed, didn't even develop the diesel locomotive. But administrators have not yet drawn the obvious conclusion. If a department is charged with a certain responsiblity, then that department will never introduce radical changes. If anyone does, someone else will do so, But so soon as he does, he will have the organization manual thrown at him. Anything he may have done will be destroyed and the task will be transferred from one who cares and knows to one who fears and opposes. At least that was my experience, and of several of my former colleagues.

This consequence is particularly unfortunate, since if my claim that the "research manager is best who manages the least" is correct, then this is the environment which will yield the largest number of outstanding results. But the consequence of illusion seven (A Rose Is a Rose Is a Rose) will be that such a manager will get a low rating and be accused of permitting "excessive duplication." Even here I can claim no priority. Albert Hirschman and Charles Lindbloom in an article "Economic Development Research and Development Policy Making: Some Converging Views," in BEHAVIORAL SCIENCE, Volume 7, 1962, pp. 211-222, explicitly recognize the benefits of a little "play" in the tight constraints placed onresearch and development efforts.

Illusion seven (A Rose Is a Rose) and eight (Procrustes' Bed) together are at the bottom of much of the ferment over "management" of research. It seems so simple to divide up the field of action and portion out support, responsibility, and facilities to each. We may even take a leaf from the military field commander, and recognizing that liaison between adjacent commands is the weak point in a battle line, we can be tolerant of a certain amount of interpenetration at the peripheries. But innovation arises from the darnedest sources. If it arises in the civilian economy or at least from non-governmental sources, we need merely create an Office of Scientific Research and

SOUND AND SUBSTANCE

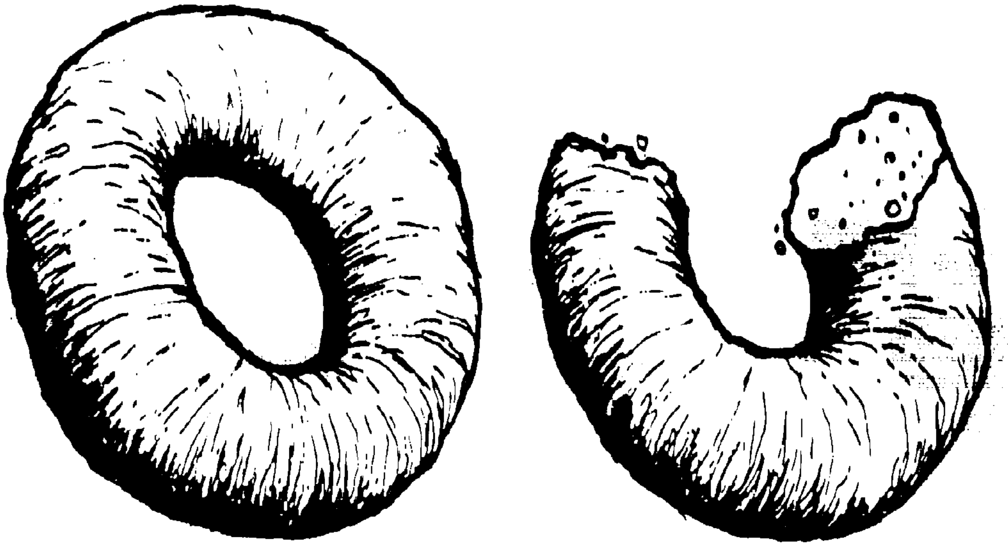
Figure 13

The financial pages of the daily papers have more than once announced the appointment of some outstanding administrator as president of a company--sometimes to rescue it from financial difficulty. The aspect I wish to draw attention to here is that in a number of such instances we learn that the former wizard has now been replaced by a new one. The criteria by which administrative competence is to be measured seem not yet always to have separated sound from substance. That is is more important to look right than to be right is the foundation of those two esteemed aspects of business-advertising and selling. It is also the foundation stone of politics (at least of the office variety). No author of fiction or movie director has the slightest uncertainty as to the great separation between appearance and reality--or as to which is better paid.

My final illusion is perhaps the most subtle. That the whole is more than the sum of the parts (Figure 14) has been most vehemently asserted by those who most abhor quantitative thinking. But the claim has been more often used as a bar to thinking than as a basis. It is the source of the claim "your vote does count"--though I think not for the reason usually given. The true reason is that, if the public would agree, it could rule, and agreement does not require that we assemble an overwhelming force, but merely that we inform ourselves. Of course, it does little good for just one voter or a few voters to do so--the answer lies in the whole (nearly) doing so. Hence it doesn't occur. Hence the principle.

This fallacy has been at the root of the argument over international trade since the Industrial Revolution. Adam Smith wrote his book to set the argument right, and while he did convince the economist, he didn't convince anyone who counted. Our individual fortunes are controlled primarily by our fate as producers, but our collective fortunes are equally involved as consumers. Since the ordinary person considers his individual (his producer) function first and foremost he penalizes himself and his fellow citizens by supporting protectionism. The most widely recognized recent large scale instance of this illusion has been exposed by John Maynard Keynes in connection with governmental fiscal and monetary policy. Fortunately, some of his resulting conclusions can be supported on valid grounds.

Public support of law enforcement "crackdowns" or of any other punitive or control measure, e.g., traffic crackdowns, compulsory



WHOLE AND PART

Figure 14

vehicle inspection, depends in large part on an assumption that it is always "the other guy" who will be affected. The risk of personally being victimized is small. But it is not zero. And the consequences (where the action is misguided) injures society as a whole--over and above the fact that, of course, while wrong actions are in progress, right actions must wait.

An exemplification closer to our topic and of vital importance lies in the field of experience. No one's own experience is "whole" till completed, whether of an individual, a firm, or a nation. Hence one cannot profit by experience as a whole while he can yet experience it. If the whole differs from the sum of the parts, it will have no effect. But the very meaning of a "rare" event is that a brief experience, possibly even that of a life time (whether of a person or of a nation) is not long enough. It has been claimed that while a fool profits by no one's experience and an ordinary man only by his own, a wise man profits by everyone's. Yet even a wise man cannot profit by experience he doesn't know about. And there are many fields in which experience is so rare that it is next to impossible to assemble it (i.e. complete it). The Constitution says the right of the people to keep and bear arms shall not be abridged. This was based on experience. But no living man can have had their experiences, and how to apply them to modern conditions is not obvious. Every teen-ager begins with much confidence and little experience and ends with more experience than confidence. That gangsters, racing drivers, entrepreneurs, dictators, explorers, and inventors base their future actions on their own past but necessarily incomplete experience is insufficiently recognized. Had all past experience been considered, many of these occupations would become deserted.

The constant advice to the young and gullible to "take a chance" in choosing a career and so secure an opportunity to win a handsome financial reward is not seen to be at one with advice to gamble on the stock market--or on a slot machine. And, of course, the prospect of a material reward (or fear of material failure) is assumed to motivate all choices in our society.

That statisticians are professionals at extracting the lessons of others' experience is recognized in a few circumscribed fields like research, development, and testing, or quality control, or acceptance sampling, but not, it seems, in opinion polling--or in safeguarding the [illegible] of a president.

My final figure enumerates these ten quantitative illusions of administrative practice. It is my belief that if administrators generally, almost universally, acquired an instinctive capacity to recognize instances of any of them on sight, despite their infinite capacity for disguise or fragmentary manifestation, then rational administration just might some day be possible.

TEQUILAP

1. Peas in a Pod
2. Profit and Loss
3. So Much the Better
4. That's Enuf, Cusp It
5. A One Dimensional View
6. Tail of the Dog
7. A Rose Is a Rose is a Rose
8. Procrustes' Bed
9. Sound and Substance
10. Whole and Part

FIGURE 15

COMBAT VEHICLE FLEET MANAGEMENT

C. J. Christianson and G. E. Cooper
Research Analysis Corporation, McLean, Virginia

Over the last six years the Operations Research Office and the Research Analysis Corporation have studied a variety of US Army vehicles from a maintenance-oriented point of view. All the studies have held the fundamental premise that to a greater extent many Army materiel decisions and programs should be based on certain mechanical properties of materiel in the real troop-machine environment. In the past the rarity of pertinent, real data too often resulted in the mechanical properties' being ignored or unrealistically appraised. It was apparent that many performance objectives were far from adequate indicators of actual achievement. ORO then and RAC now have sought to introduce realistic measures of mechanical effects into managerial decision processes. Of necessity this management research mission had to be confined to limited numbers and types of equipment. In order to determine meaningful equipment policies, it is necessary to consider troop-performance data in combination with a variety of monetary costs and with relative obsolescence. Neither dollars, nor obsolescence, nor mechanical performance alone can be expected to give infallible guidance to the necessary consideration of equipment management. Occasionally technological breakthroughs do occur, and sometimes a particular equipment model does develop a rash of breakdowns. However, in the long run the Army has to program much of its inventory between the relatively gradual changes in the designs for new production and the mechanical aging of previously manufactured equipment.

The most current study of Army vehicles has been selected as a general example of the ways in which RAC has contributed both greater qualitative understanding and improved numerical assessment techniques to the solution of important materiel management problems. Certain of the results have been disguised for relatively open presentation; however, wherever possible, numerical examples have been kept along the scales of their true values. Indeed, there is often a need to force experts to recognize the order of magnitude of the effects with which they profess intimate qualitative acquaintance.

1—The Research Analysis Corporation recently completed a comprehensive, multi-stage analysis of three different types of tracked vehicles operated and maintained in US Army combat units in Europe. A main

objective of such study was to determine in-use lives of tanks, armored personnel carriers, and recovery vehicles. In addition to fulfilling its principal objective, the study program provided a variety of by-products essential to successful management of combat vehicle fleets. Measurement of fleet wearout, establishment of repair capacity requirements, projection of budget needs, assessment of combat readiness, and determination of fleet replacement factors (procurement requirements) are all management responsibilities that cannot be successfully fulfilled without basic information and analytical results of the type provided by RAC.

PROBLEM

COMBAT VEHICLE FLEET MANAGEMENT

Utilization
Component Replacement Forecasting
Readiness

Study conducted for the U. S. Army
by the Research Analysis Corporation

Figure 1

Inasmuch as the same analytical procedures were applied by the study to all three vehicle types, little generality is sacrificed if most further discussion is limited to just one vehicle--the tank. The determination of in-use lives is treated here as only a part of the general problem of tank fleet measurement. Within a manageable framework equipment life is interrelated with factors of utilization, component replacement forecasting and costing, and materiel readiness, to name only a few. The factors cannot be divided into one list of independent and another of dependent variables. They can be functionally related, but any adopted conventions of dependence and independence are likely to be unrealistic statements of causes and effects.

In all RAC equipment studies the starting point of meaningful concepts has always been a body of carefully collected empirical data. The hypotheses preceding observation have almost always had to be so modified under the light of experience that it now proves preferable to avoid prejudicing

early theories. The logical elegance of classical scientific method has to be modified in favor of diffusely related series of observations resulting in near-saturation with both pertinent and extraneous data. Data processing and analysis then have to be not so much testing of theory as they must be sorting of the relevant from the irrelevant. Of necessity the rules of relevancy must be continually modified as more and more data are digested. And discouragingly it often becomes essential to completely reprocess earlier work.

Fig. 2—The body of empirical data used for the most current study accrued over the complete history sample is now shown. Maintenance events that occurred against all these vehicles were recorded by vehicle number, age, and mileage. The column on the right gives ample evidence of the fact that large numbers of combat vehicles are used extensively in peacetime. Peacetime is unquestionably a time of appreciable materiel consumption. That the severity of peacetime demands is still not universally recognized in all related civilian and military agencies continues to provide one of the major obstacles to completely successful combat vehicle management.

SAMPLE

		Average		
	No. of vehicles	Months observed	Miles observed	Usage miles per mo.
Tank	640	21	2739	130
Armored Personnel Carrier	708	18	2655	148
Tank Recovery Vehicle	83	19	1485	78

Figure 2

Fig. 3—One of the fundamental operations performed on the body of empirical data is the construction of mileage-dependent parts costs. The average direct costs for tank repair parts are shown for each successive 500-mile increment to 4500 miles. Track and engine replacements account for most of the dollar consumption. One of the most important features of parts consumption is the strong increase with mileage beyond 1500 miles.

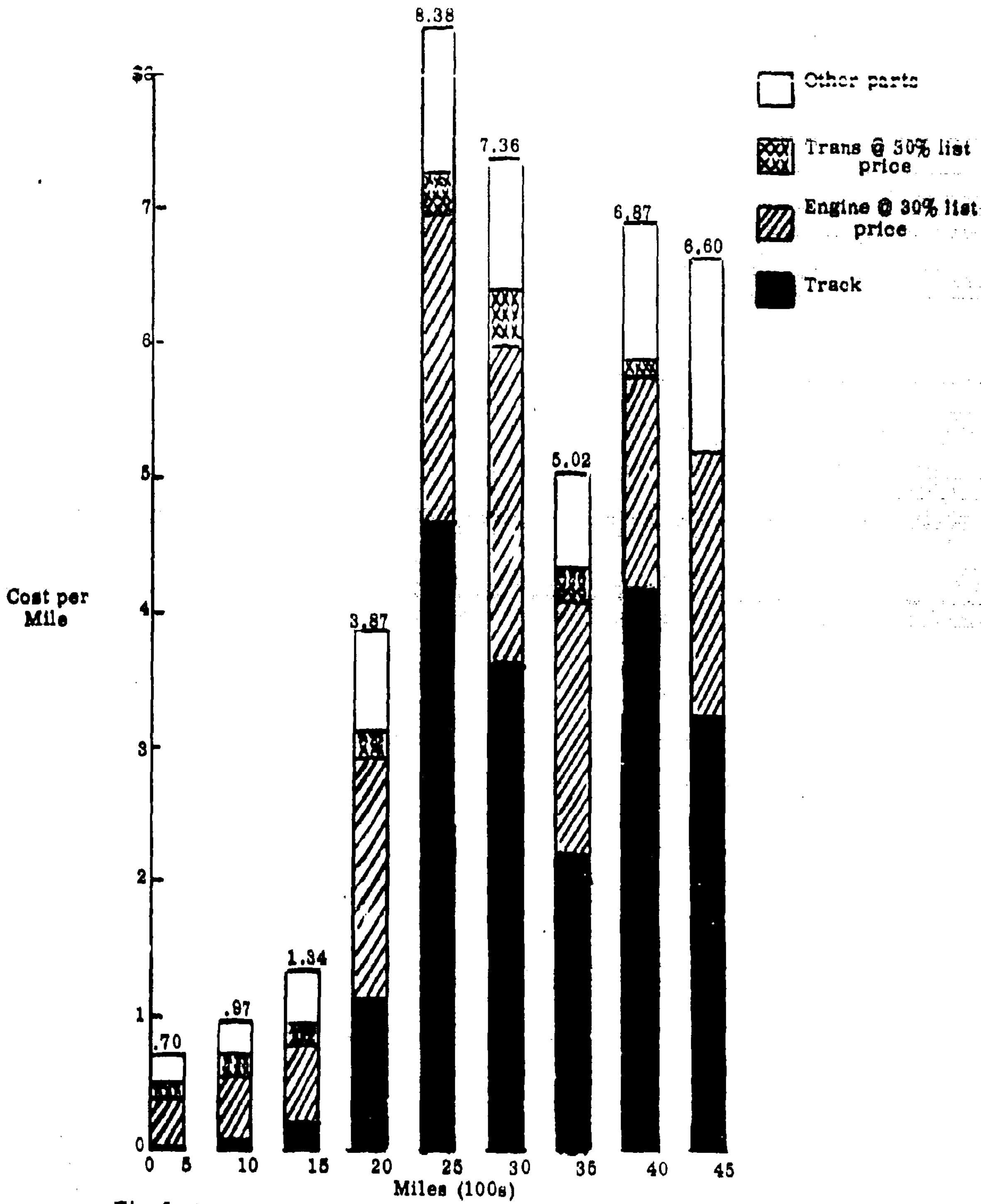


Fig. 3– Average Cost Per Tank by 500-Mile Increments
Total Miles Observed 1.75 Million

Through 1500 miles parts expense averages very close to $1 per mile. Beyond 1500 miles both track and engines enter pronounced replacement phases. Modal track replacement occurs at about 2200 miles. The dollar effect of this mode is to drive parts costs above $8 per mile in the 2000- to 2500-mile interval. Beyond 2500 miles there is a definite decline in expense, but a second generation of track replacement beyond 3500 miles forces the costs up again. From a separate examination of the same basic data it was estimated that, if all subsequent installations of parts and assemblies provided lives similar to those of the originally installed parts, an equilibrium cost would just exceed $6 per mile for a list of the principal mobility-affecting parts. Inclusion of weapon and fire control costs would add to the $6 per mile figure. The changes in parts consumption with mileage obviously have tremendous impact of the provision and budgeting of parts and maintenance support. If tanks are operated at 1500 miles per year, support during the first year of tank life be only one-sixth that required during an equilibrium year. Too much support planning is still dependent on an assumption that each new year is going to be like the last one. The importance of making predictions based on analyses of trends is revealed by data of the cost type.

Fig. 4—For the same tank sample, the purely historical cost average to nearly two years was $3.75 per mile. The truth of such history does not make the past a guaranteed base for prediction of identical futures. Vehicle support must be programmed in calendar time. In order to provide accurate support it is necessary to know both the basic consumption rates per mile and the mileages to be covered during given calendar intervals. Again in the case of the tank sample, adjustment of the average mileage cost to the average use rate yields monthly costs of $489 per tank per month. At the same rate of travel, but using tanks with much higher mileages, corresponding expenses would exceed $780 per tank per month.

Figure 4
PARTS COSTS FOR TANKS
(Usage: 130 miles per month)

	Cost per mile	Cost per month
Engines*	$1.19	$155
Transmissions*	.19	25
Other Parts	.32	42
Track--amortized at $2.05 per mile	2.05	267
Total	$3.75	$489

*30 percent of list price.

Fig. 5—The occurrences of replacement events are more fundamental than their associated costs. One of the advantages of the format employed for collection of the basic history data is that mortality statistics for individual parts types can be determined by straightforward actuarial calculation. For example, the comulative engine replacement experience is shown for the total tank sample. The accumulation of replacements increases non-linearly to just beyond 2000 miles, but then the engine activity becomes almost uniform.

Fig. 6—The replacement of original engines was interpreted as having been governed by an underlying distribution like that labelled first. That distribution is represented by two distinct phases. The first, extending to just beyond 1000 miles, averages only about one-half percent replacement per 100 miles. This first phase corresponds closely to the typical early "debugging" period so common to many kinds of equipment. However, unlike the frequently discussed "bath-tub" effect, tank engines do not then experience a phase of reduced replacement activity. Rather, the tank engines immediately being a second phase with sharp increase in the replacement rate. The second phase corresponds to what one would expect with entry into a wear-like mode of engine mortality. The presented distribution of mileages to replacement has a mean of 3700 miles.

If it is assumed that all succeeding engine installations will result in lives like those of the original engines, second and third replacements will occur as shown by the distributions labelled 2d and 3d in the figure. Such higher order replacements have been determined by taking repeated convolutions of the first distribution. The sum of replacements of all orders is often described as the renewal density. It represents the instantaneous replacement rate disregarding the order of a replacement. A pure coincidence of the chosen first replacement distribution is that its corresponding renewal density approaches equilibrium quickly but smoothly. In fact, by 3000 miles the renewal density is almost equal to its equilibrium value. The absence of strong oscillations in replacement activity is usually of great advantage in the prediction of maintenance support requirements. Somewhat later comments will be made by way of explanation of real effects that tend to drive engine replacement activity above the so-called equilibrium presented here.

Fig. 7—Conversion of a renewal density of corresponding budget costs is a relatively easy matter. Multiplication of the renewal density by the unit cost of a replacement yields the desired measure of expense. As an example of the type of results derived by such a processing operation,

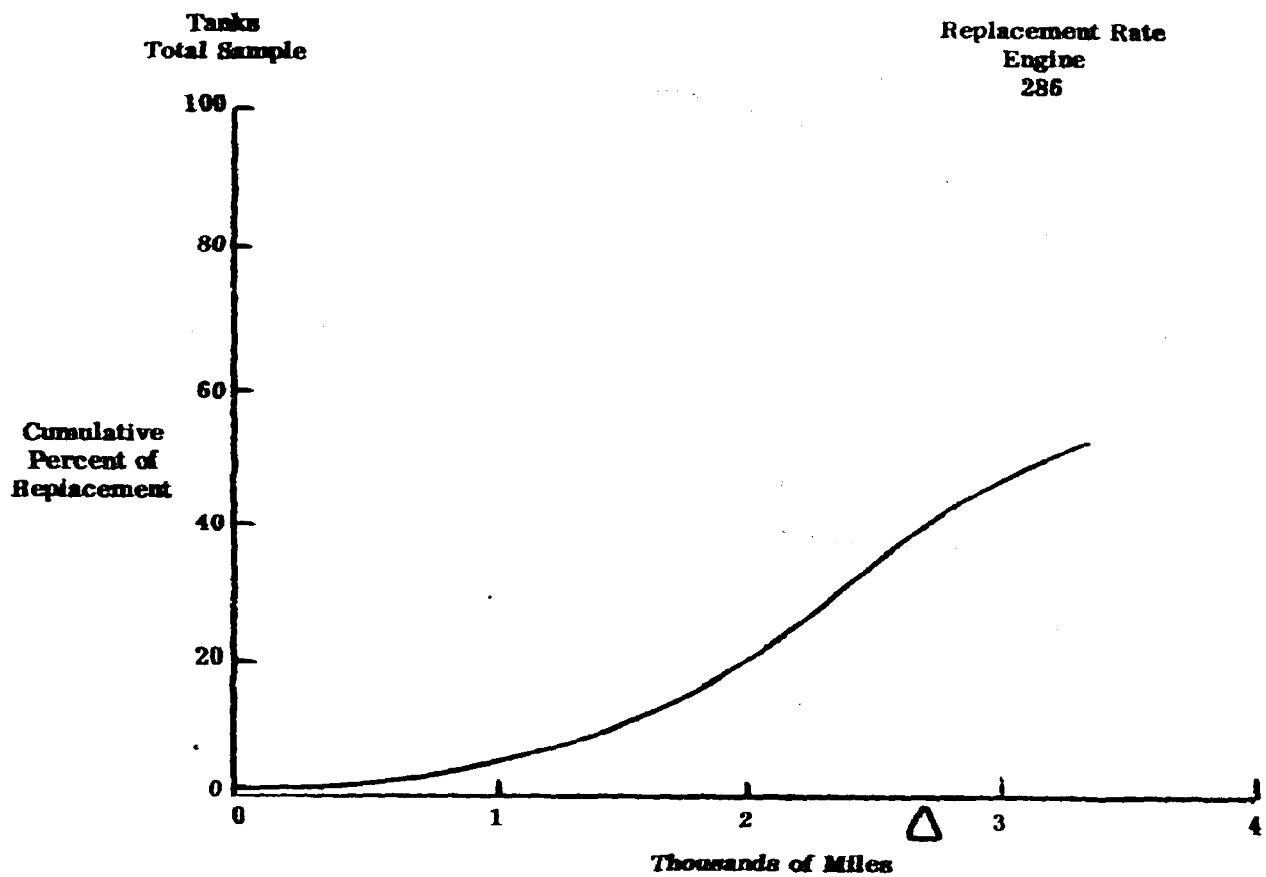
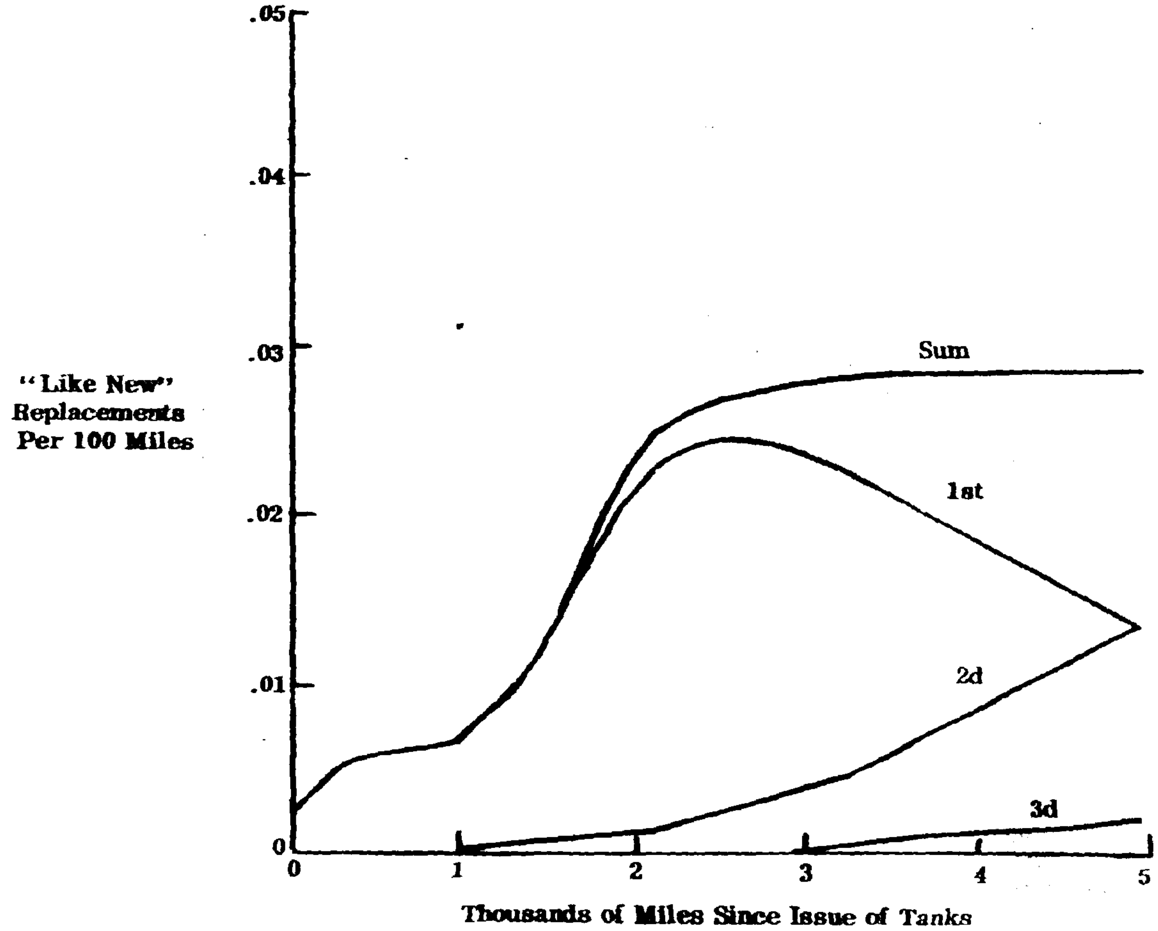

Tanks
Total Sample
Replacement Rate
Engine
286
Cumulative
Percent of
Replacement
100
80
60
40
20
0
0
1
2
3
4
Thousands of Miles

Fig. 5

Fig. 6–Tank Engines at 130 Miles/Month



a cost rate dependent on mileage is shown here for engine and transmission replacement activity. Such information represents interpretation and extrapolation of the originally collected empirical data. Transmissions have lower unit prices and longer lives than do tank engines. This double advantage to the transmissions makes their costs per mile considerably lower than that of the engines. Note that the total expense for these two assemblies accounts for nearly one-half the previously mentioned equilibrium cost rate of roughly $6 per mile.

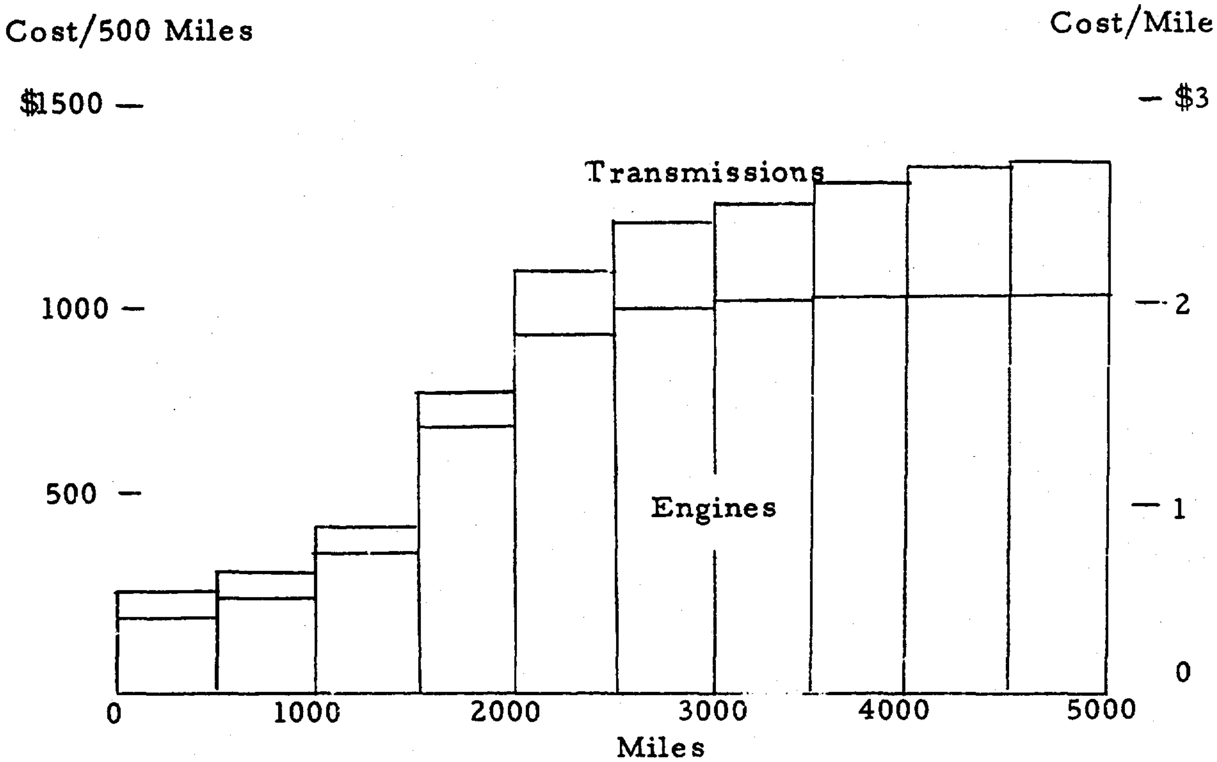


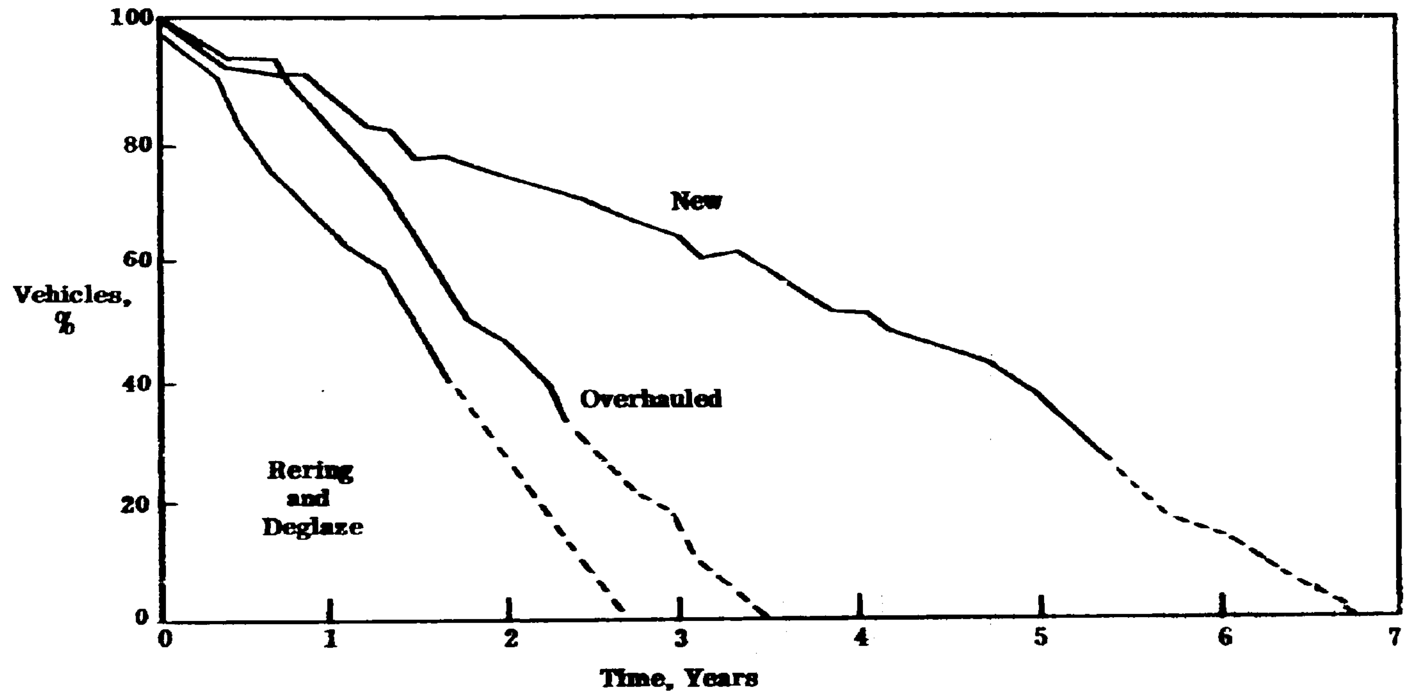
Figure 7--Tank Major Assembly Replacement Cost "Like New," 30% of List at 130 Miles/Month

Reality usually has a way of muddying the clear waters of analysis. The preceding engine and transmission predictions are all founded on an assumption of performance from replacement assemblies like that of the originals. Unfortunately very little data have accrued about the performance of replacement engines in the model tank most recently studied.

Fig. 8—However, the experiences of a preceding model tank and its engines are not ignorable. There survivor curves are shown in the projected figure. All these were determined for the other tank. That tank's original engines lasted as shown by the line labelled "new." Engines that were overhauled by a depot maintenance facility survived only as long as shown by the curve labelled "overhauled." Other engines were repaired by a mobile team (fourth echelon in the US Army). The lives of engines repaired by that team were even shorter and are represented by the line labelled "re-ring and de-glaze." After several years virtually all the engines that are installed in used tanks are repaired ones. The possibility must be recognized that equilibrium engine replacement activity may run twice as high with repaired engines as that determined from the originally performing assemblies.

Fig. 9—For the current model tanks, second engine replacements have been running somewhat higher than expected. Only about 10 percent of the second engines are known to have been repaired ones. The rest are presumed to have been issued from storage in unused condition. If the second engines were to do as well as the originals did only 4 percent of the tanks would experience a second replacement by 3000 miles. The predicted "like-new" experience is represented by the lower solid line in the accompanying figure. Empirical experience is shown by the lower dotted line and runs just over 8 percent replacement to 3000 miles. The real replacement mechanism seems to be one that is only partly renewal-like. Tank age regardless of assembly age appears to provide an important component of the probability of replacement. By now listeners are probably wondering how the pure renewal model can be of any use of management if so many of reality's disturbing influences tend to raise activity much higher. The renewal model provides a valuable reference point. First of all, the renewal estimates have already revealed that many support programs are set even lower than these predictions. The pure renewals usually specify the best that can be expected. If support is too low and is not geared to satisfy the demands of the best possible, it should be raised immediately to at least the level consistent with a

Fig. 8– Survival Rate, Tank Engines



renewal prediction. And second, the renewal estimates provide the basis of comparison of performance of repaired, stored, or modified assemblies with original quality. Too often statements are made that some component is good or bad without expressing goodness or badness relative to a real mechanical standard. The usual comparisons with paper standards have caused more confusion than they have eliminated. Even original engines do not compare favorably with their paper standards by which they were designed, built, and procured for Army use. Although performance of materiel in the hands of troops cannot be the ideal, absolute standard of reference, it remains the best, most meaningful basis of comparison at the present time.

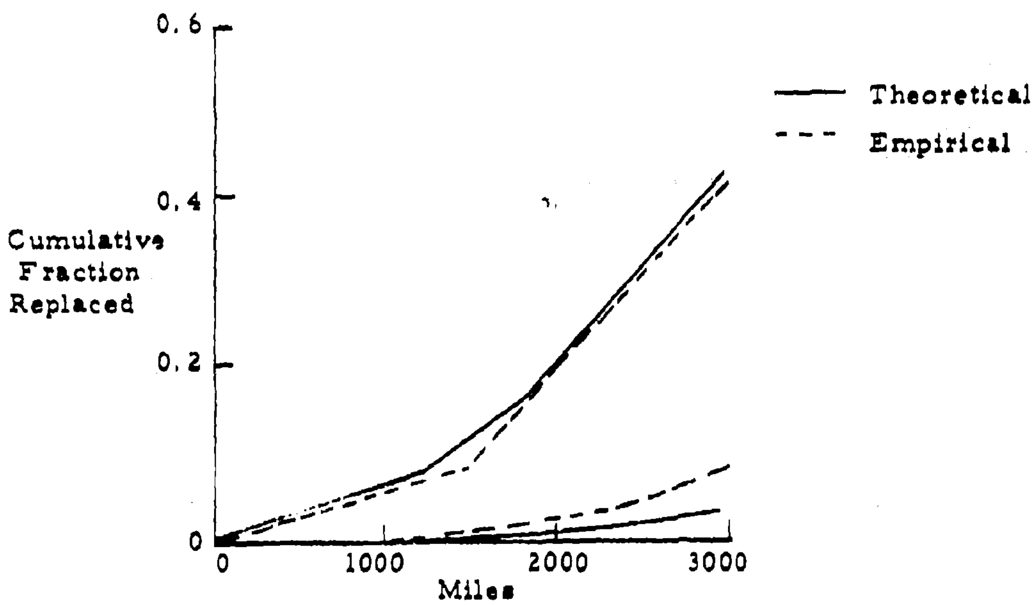


Figure 9--Comparison of History with Expected Like-New Renewals of Engines

Fig. 10—The replacement distributions of most parts and assemblies are widely distributed about their means. A particularly important example of a part that does not have this property is tank track. The lighter line in the projected figure shows the cumulative replacement experience for sets of tank track. The S-shaped curve represents the most compact distribution of replacements discovered in the whole series of studies of ground vehicles. A complete set of track is expensive. At the same time it has the shortest life among all the high-unit-priced parts of a tank. Conversion of life and unit price to costs per mile yields an estimate of \$2.05 per mile for tank track. Although engines and transmissions cost more than track, their lives are sufficiently long to make their mileage costs lower than that of track. Track expense accounts for roughly one-third the predicted equilibrium parts costs for the mobility-affecting systems of tanks. A cumulative curve for second replacements is also shown. RAC pointed out that the early climb of the second curve was in part attributable to the use of much shorter-lived rebuilt track.

Fig. 11—Much effort has been directed toward interpretation of the empirical history relative to notions of materiel readiness. No one definition has proved satisfactory as a full description of readiness, but several less general notions have proved particularly useful. One concept that has been extremely helpful is that of "equipment availability potential." An item of equipment is considered to be assignable to one of several states of serviceability. For example, consider the four states shown. An item can be serviceable or it can be in one of three (or more) states of unserviceability. That item can undergo transitions from one state to another with probabilities associable with each particular type of transition. The "k's" can be associated with breakdown, and the "λ's" can be related to correction or repair times. The transition probabilities may depend on ages, mileages, generation numbers, or other factors. Too often it is not possible to detail all the inter-relations. In fact much of the time the greatest utility arises from considering a two-state model; that is, one state of serviceability and only one state of unserviceability. The instantaneous "availability potential" is defined as the probability of being in the serviceable state if transition probabilities were to permanently retain their current values. When the transition probabilities change sufficiently slowly, the availability potential gives a suitable measure of actual current serviceability or availability. Constant levels of availability potential may be represented by hyperbolas on a "k-λ" plane.

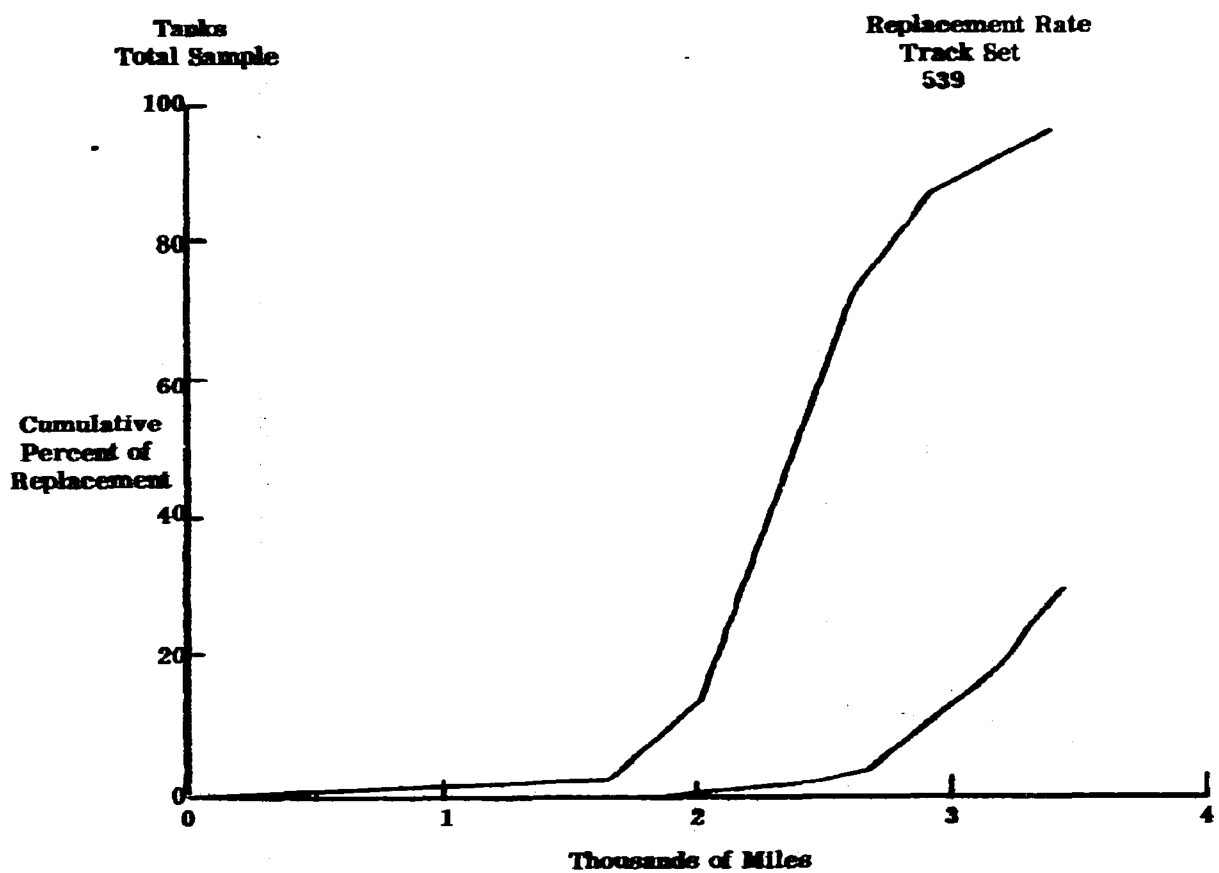
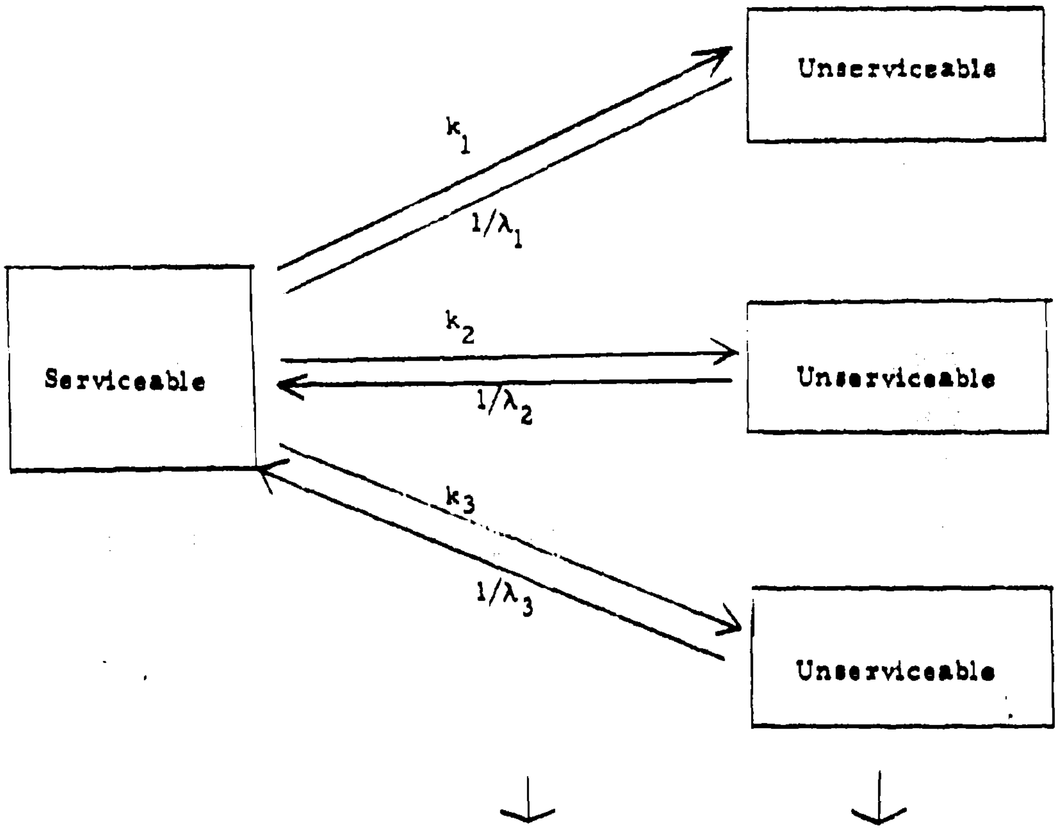

Tanks
Total Sample
Replacement Rate
Track Set
539
Cumulative
Percent of
Replacement
100
80
60
40
20
0
0
1
2
3
4
Thousands of Miles

Fig. 10

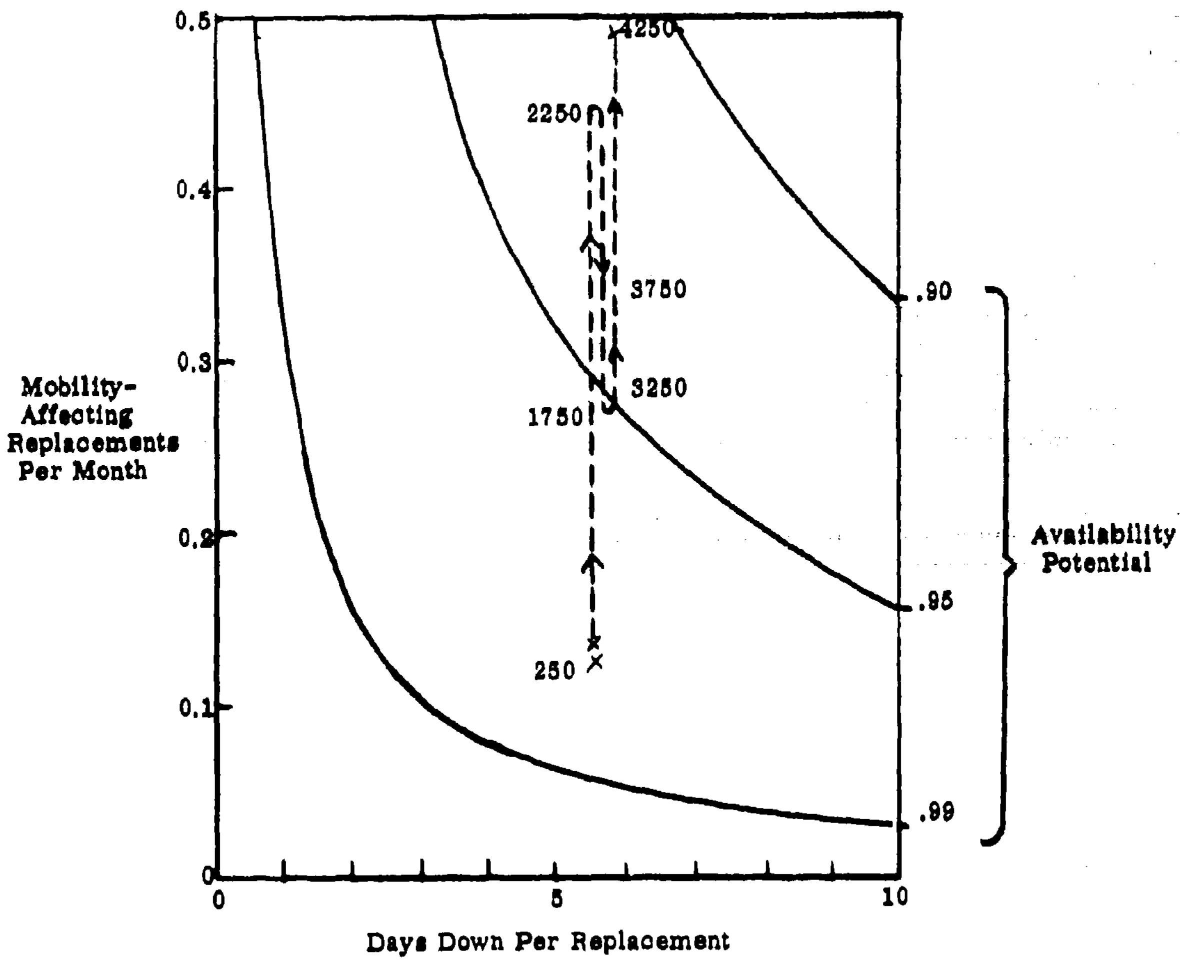


Concept
Availability Potential

Figure 11

Fig. 12—Analysis of the mobility-affecting parts replacement data from the same tank history sample led to the mileage dependent availability path shown. During the entire history period the average response time stayed very close to 5.7 days per replacement job. However, because the rate of replacements per mile was changing, the availability at 125 miles per tank per month had to change as shown. During their early lives the studied tanks were operated with close to 0.98 availability potential. As they accumulated additional mileage, the tanks lost

Fig. 12--Tank Availability Potential
At 125 Miles per Month



availability. During the period of most track replacements, the availability dropped to about 0.95. Then followed a period of some improvement; the availability climbed to 0.95. However, beyond 3250 miles availability again dropped and reached about 0.91 in the interval 4000 to 4500 miles.

Fleet availability need never be considered an unmanageable aspect of operation and maintenance. Rather it is only a result of several directly manageable factors. Product quality, support response time, and rate of equipment use all affect equipment readiness.

Fig. 13—A great deal of attention is usually given to the supposed or predicted differences in product quality among different models of equipment. The normal approach of salesmanship is to promise that some new model will provide mechanical advantages far beyond the cabilities of its predecessor. Too often the demonstrated comparisons examine an unused, new model and an over-used, old model. The RAC studies have discovered that many models of different generations appear so different only if they are examined at different stages in their lives. At the same ages different models of tanks possess greater similarity with respect to parts replacement rates than do tanks of a single model at different ages. Judicious utilization of a particular tank model can increase overall mechanical capability more than can a transition of models amid a less carefully designed program of vehicle use.

As an example far less pronounced than reality, consider this figure. Suppose that at a uniform rate of use some tank model has the renewal density shown with respect to time. That density increases smoothly. A change of utilization can have a three-fold effect with respect to time. An increase in use, in effect, squeezes time by having the higher accumulated mileages occur that much sooner in time. Thus, in a given month the replacements per mile are higher. Then because more miles are traveled during that month, the replacements during a given month are given a second boost. The third effect may be to somewhat alter the mileage lives depending on the rate of use. It is not illogical to consider the possibility of increasing or decreasing mileage lives depending on the type of component involved. Actually it is usually sufficient to suppose that the occurrence of events per mile depends on the accumulated mileage but not on the rate of use. This assumption is nearly correct within the mileage ranges normally encountered and corresponds to having(u) equal to 1.0.

Fig. 13—Three-Fold Utilization Effect

$$h(t,u) = \left(\frac{u}{u_o}\right)\phi(u)\, h\left[\left(\frac{u}{u_o}\right)\phi(u)\, t,\ u_o\right]$$

I II III

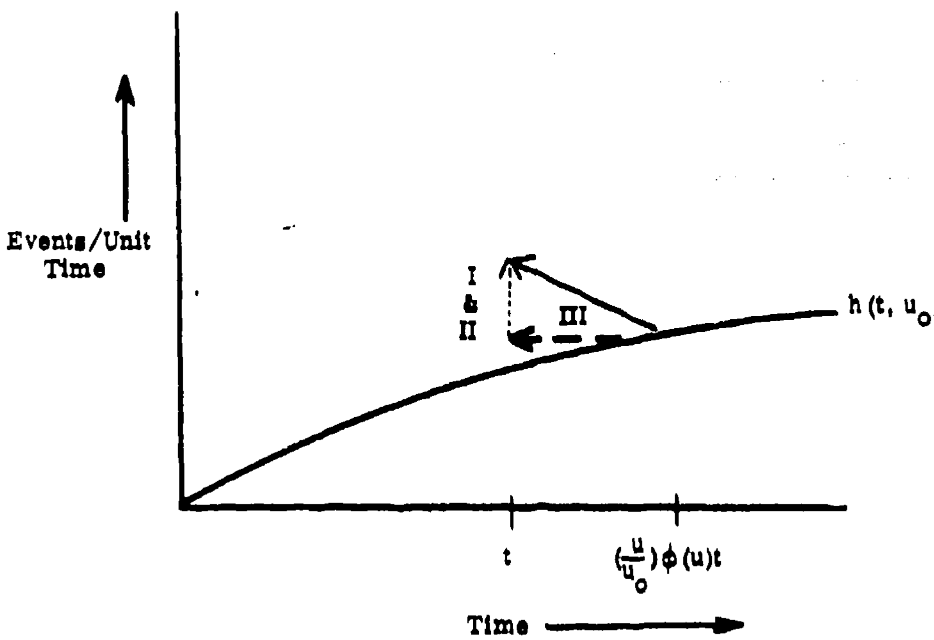


Fig. 14—Now consider an example from the real life of the studied tanks. In the accompanying illustration the calculated renewal density for engine replacements is shown along a time base of tank use at 130 miles per tank per month. At that rate of use engine replacements at 20 months amount to about 4.6 per 100 tanks per month. Now consider what would happen if use were increased 50 percent to 195 miles per tank per month. At 20 months the faster tanks would have the same total mileage as do the slower tanks at 30 months. At that mileage the engine replacement rate per mile is higher than at the lower mileage. At the same time the tanks are going 1.5 times as many miles per month. The net result is that at 20 months tanks going only 1.5 times as fast may be expected to experience over 1.8 times as many replacements.

Next consider the prospect of having operated those same tanks at only one-half of 130 or 65 miles per tank per month. At 20 months the slower tanks have accumulated only as many miles as had the 130-mile-per-month tanks at 10 months. Thus at 20 months the slower tanks experience engine replacement at a much lower rate per mile, and because the slower tanks cover only half as many miles per month, their engine activity is even lower. In fact, at 20 months the reduction of use by one-half results in only 0.17 times as many engine replacements.

The magnitudes of change with respect to time that can be effected by utilization control are obviously greater than many of the differences asserted to exist among different models. It should also be obvious that the utilization impact extends to materiel readiness, assembly repair, assembly floats, maintenance allocations, and so on throughout much of fleet management.

Fig. 15—A nomogram was constructed as an illustration of the utilization effect on major assembly maintenance activity. The nomogram expresses the interdependence among mileage replacement rates, equipment utilization, time replacement rates, durations of repair pipelines, and assembly float requirements.

For example, consider an assembly that is being replaced at a rate of one per 3700 miles (reference the point along the lower left scale). A vertical trace to roughly 130 miles per month reveals that 1000 tanks experience about 38 replacements of that assembly per month. If three months are required between the removal of the assembly to the time that it is repaired and available for re-use, it is necessary to keep just

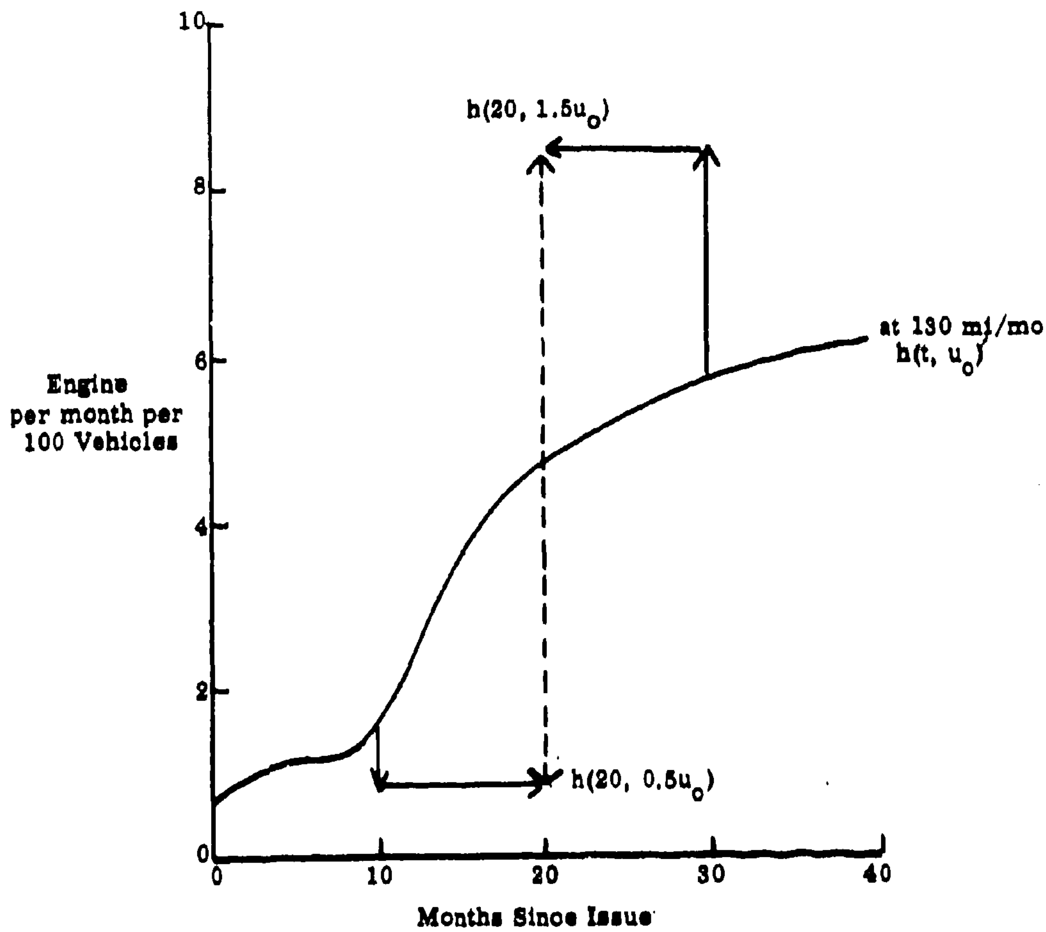


Fig. 14—Direct Effect on Assembly Replacements of 50 Percent Changes in Vehicle Use

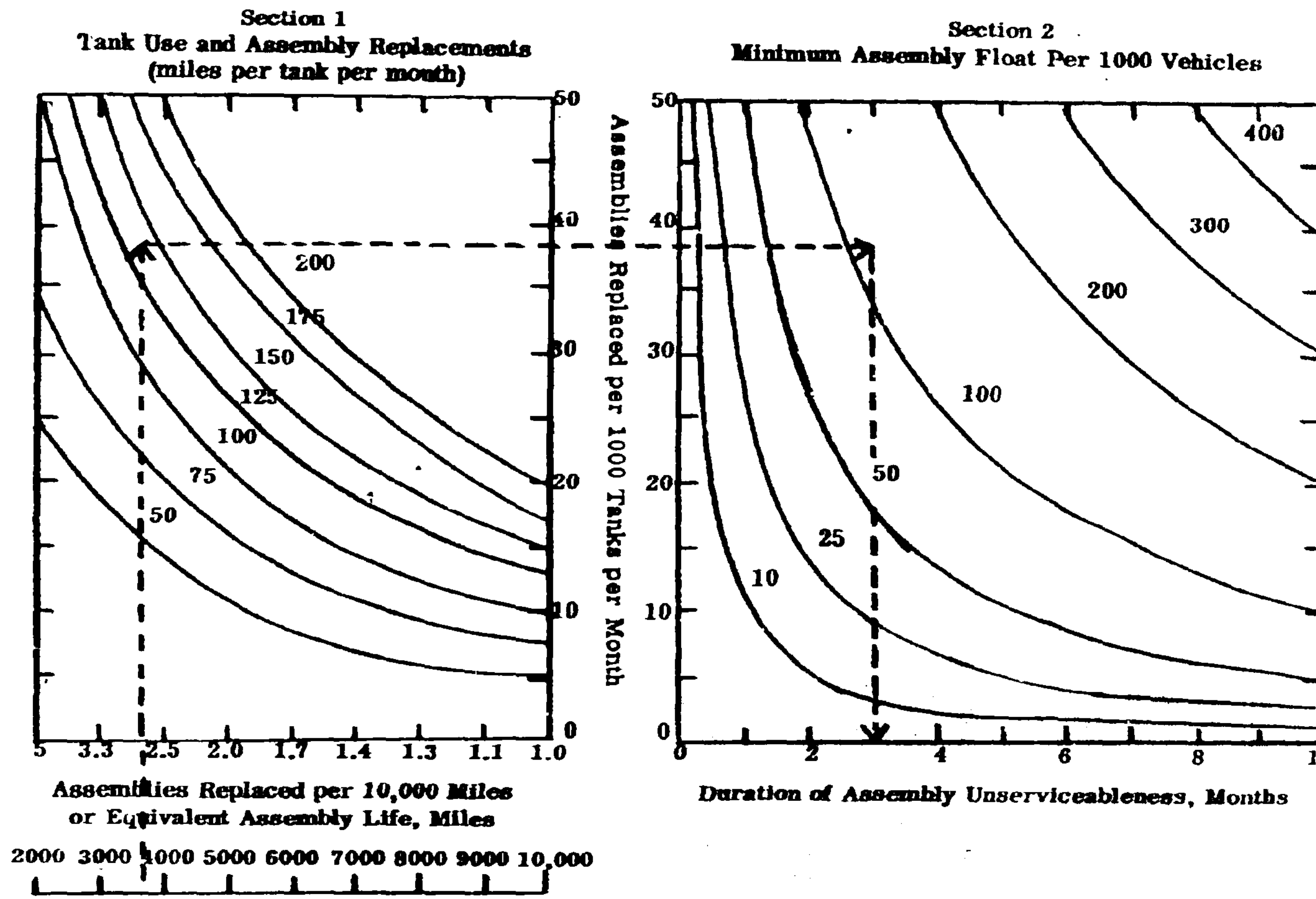


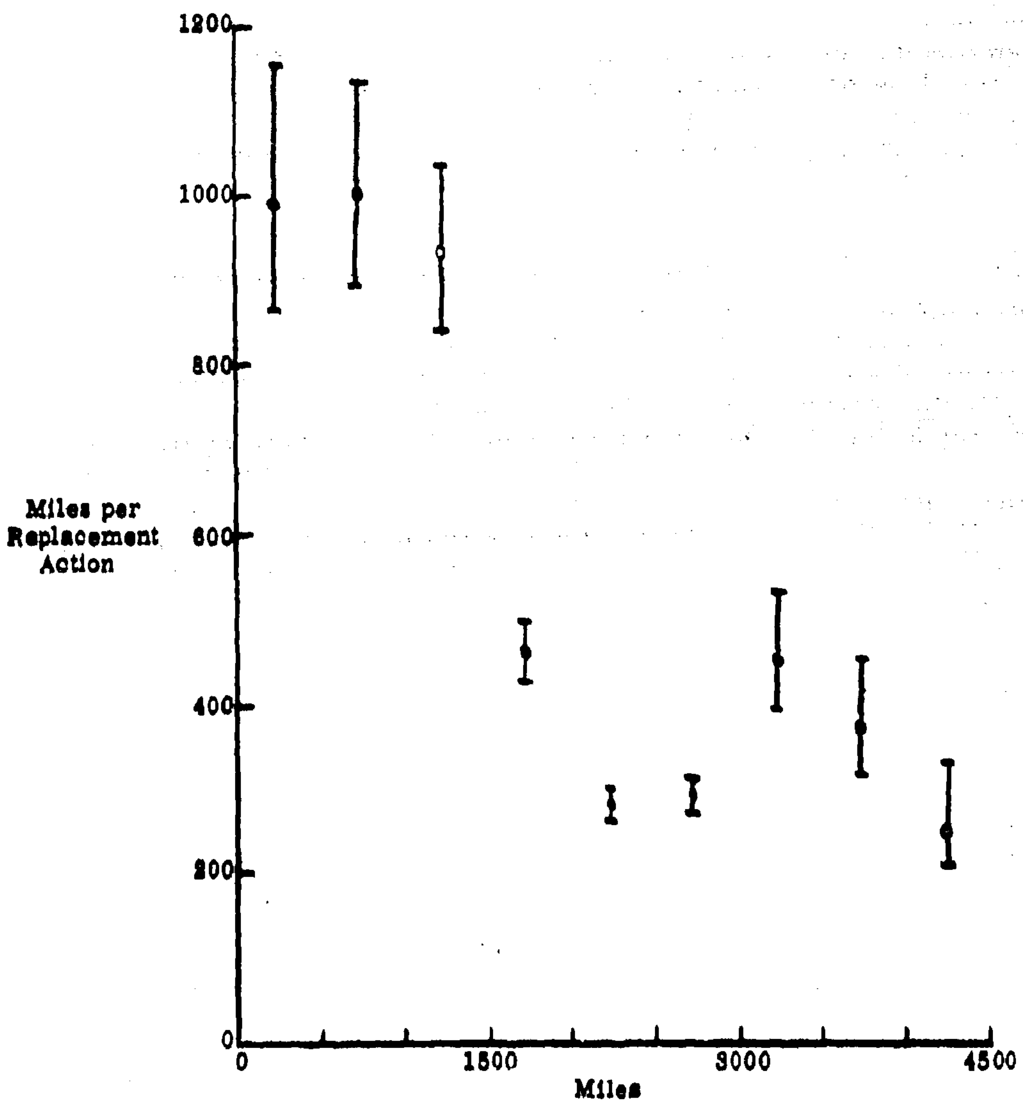
Fig. 15—Relation Among Assembly Life, Rate of Use, Duration of Unserviceableness and Minimum Assembly Float

over 100 of those assemblies in the assembly float or pipeline. Efficient exploitation of combat vehicle resources demands that close, continuous attention be given to all the quantities described in the nomogram. At a given time the fleet generates demands for assembly replacements in a way that depends on both the assembly quality and the fleet use. The duration of the pipeline depends on heavy maintenance programs, stocks of parts, and the geographic location of facilities. A trans-Atlantic pipeline by surface transport can easily run to many months and result in gigantic increases in the required float size.

Fig. 16—The aggregate of mobility-affecting parts replacements provides a basic indicator of what to expect in the way of vehicle performance. Such data were already used to determine the mileage dependent availability potentials at a given rate of use. Perhaps a more fundamental way of viewing the replacement activity is to consider the average miles per parts replacement action over a range of mileages. Such information is provided in the projected figure. To 1500 miles the studied tanks performed with about one replacement action per 1000 miles. Beyond 1500 miles the performance dropped rapidly with the incidence of a great deal of track replacement activity. Improvement occurred in the 3000- to 3500-mile range, but then a decline again appeared. From the detailed basic data it was estimated that the trend would eventually lead to an equilibrium activity close to 165 miles per replacement job for the mobility-affection parts. This level is based on the assumption that all installed parts and components last as well as did the originals.

Fig. 17—The figure now presented provides an example of the effect of using repaired assemblies as replacements in older tanks. Tank A represents the model studied most currently. The tanks down for engine or transmission replacement are shown for Models A and B when all replacement assemblies perform as well as did their originals. Model B was actually studied several years ago, and considerable data were collected for it during periods when it did receive repaired assemblies. Model B was operated at a much lower rate of travel, but translation of its major assembly experiences to the same tank use rate as that of Model A led to the much steeper line. In other words, had the older model tanks been operated at 130 miles per tank per month, at 40 months they would have most likely experienced about a 5-percent deadline rate for engine or transmission replacements assuming enough assemblies would have been available from the repair facilities. Model A was not observed much beyond 20 months

Fig. 16–Tank Miles per Mobility-Affecting Replacement Action



and hence had not had an opportunity to experience use with mostly repaired engines. The similarity between the predicted behavior of both Models A and B with all-new assemblies is cause to suspect that Model A might then do as poorly as Model B when mostly repaired assemblies are provided it. A very likely consequence of the observed trend of performance is that fleet users will probably, quietly reduce their level of tank use to one allowing more relaxed maintenance support.

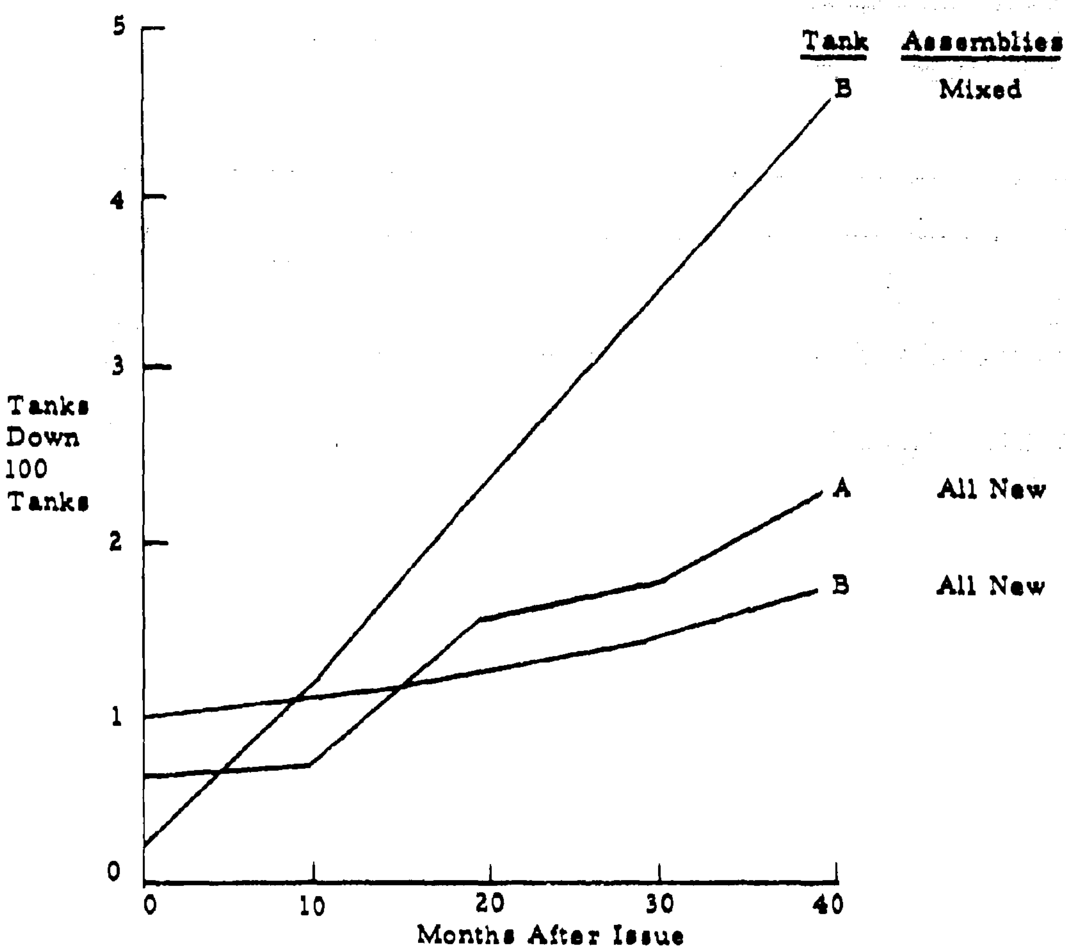


Figure 17--Tanks Down for Engine and/or Transmission Replacement at 130 Miles per Month

Fig. 18—Equipment availability is only a partial measure of materiel readiness for combat. In general availability can always be increased by decreasing the rate of use. Such a phenomenon seems to occur very often. A matter of equal importance to readiness is the performance to be expected from any equipment that is available. Vast segments of the US Army's inventory face a very severe dual requirement. Such equipment is used in extensive peacetime training programs. The equipment has to be available not only for training but also for any emergency deployment to combat. In order to survive if combat should arise, it is necessary that the equipment continues to possess an adequate of residual or combat life. The preceding examples from tank life give sufficient evidence of a probable loss of residual life as the equipment is used and ages. The data of mobility-affecting parts replacement activity were further translated into a measure of what would be expected in the way of tank endurance in the event of an unexpected 50-mile march over hard-surfaced roads. In order to score a success, the tanks subjected to this hypothetical test must be available to start such a march and then complete it without a mobility-affecting deficiency. The march was made short for several reasons. The principal reason was that over several years, the observation of several scheduled marches revealed that tanks experienced most of any deficiences to about 100 miles during the first 50 of those miles. Hence a success to 50 miles is good assurance of success to somewhat more than 100 miles. Too many so-called readiness tests are not previously unannounced. On announced exercises units very often are able to deploy all their vehicles initially. Were a deployment requirement to be issued at some other time the results would be likely to be far different.

For the hypothetical march measure, the mileage-incidence of mobility-affecting replacement actions was assumed to follow the trend developed from the main body of tank history data. In the particular chart shown the training requirement was taken to be about 125 miles per month. Over a period of many years tanks would continue to accumulate mileages, lose availability, and do less well against the unexpected march. The 10,000-mile line is just one rough approximation of what an equilibrium tank might achieve. The march model has been given considerable elaboration with study devoted to the implications of various functional effects for different utilization dependencies. The requirement for uncomplicated visual presentation has led to the reduction of all results to formats very similar to that shown.

Fig. 18—Effects of Age on Hypothetical Performance Tanks on a 50-Mile March

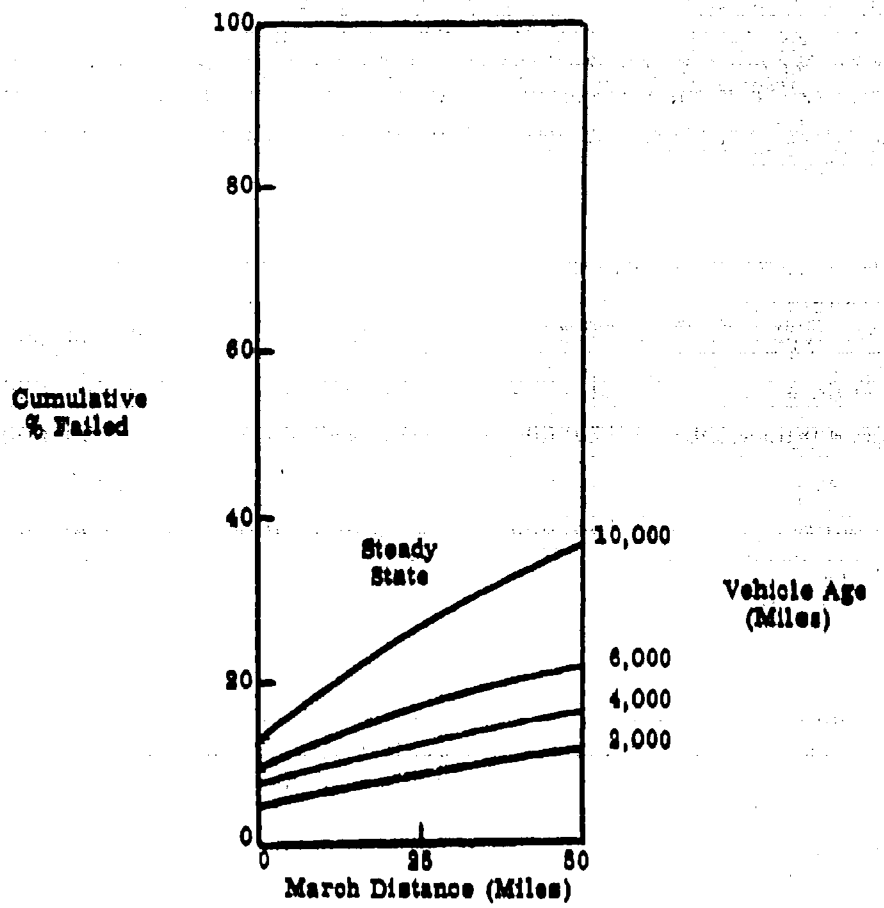


Fig. 19— As part of the conclusion of this general survey of RAC's management assisting activities in the area of combat vehicles, it is appropriate to reintroduce consideration of parts support costs. At 125 miles per tank per month the yearly support of a fleet of 1000 tanks of the model studied requires the funds shown in the accompanying table. During the first year the fleet consumes about $1.5 million worth of parts and assemblies. By the third year over $10 million worth are being consumed. Track replacement by far comprises the biggest single slice of the total bill. In the third year the "other parts" account for only about 16 percent of the total cost, but they involve a great variety of different kinds and numbers of repair parts.

ANNUAL EXPENSE--1000 TANKS

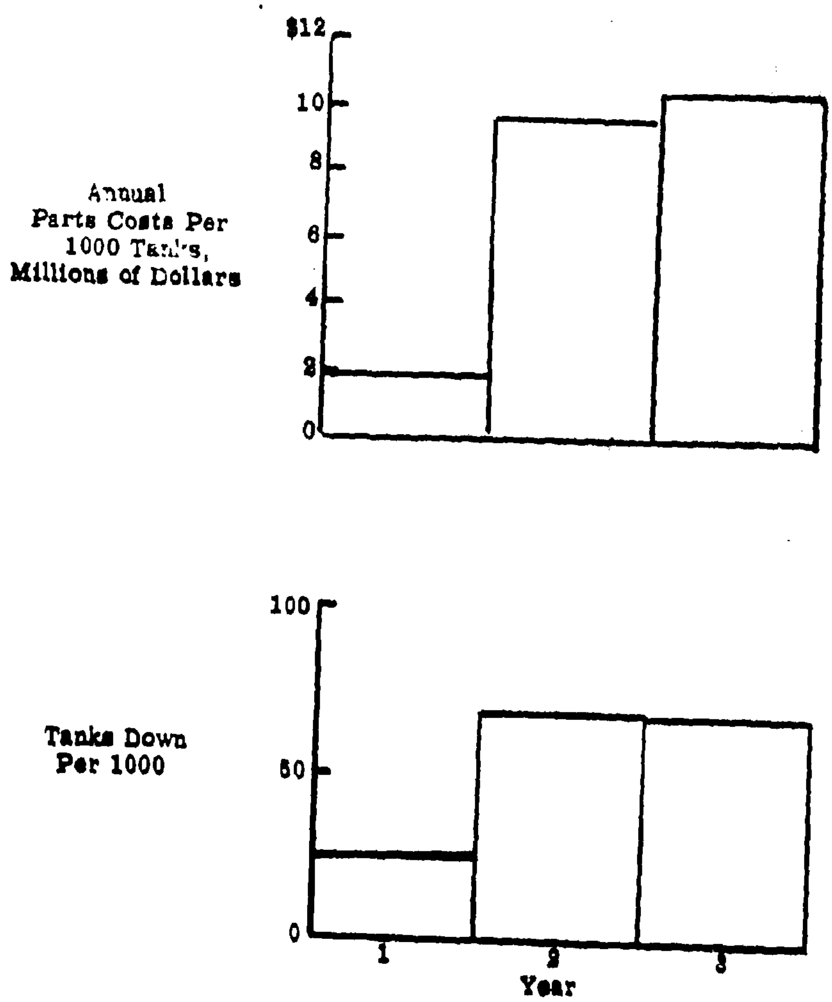
Year Mileage	1 0-1500	2 1500-3000	3 3000-4500
Engine	$ 740,000	$2,540,000	$ 3,000,000
Transmission	170,000	430,000	830,000
Track	500,000	5,130,000	4,900,000
Other	460,000	1,400,000	1,600,000
Sum	$1,870,000	$9,500,000	$10,330,000

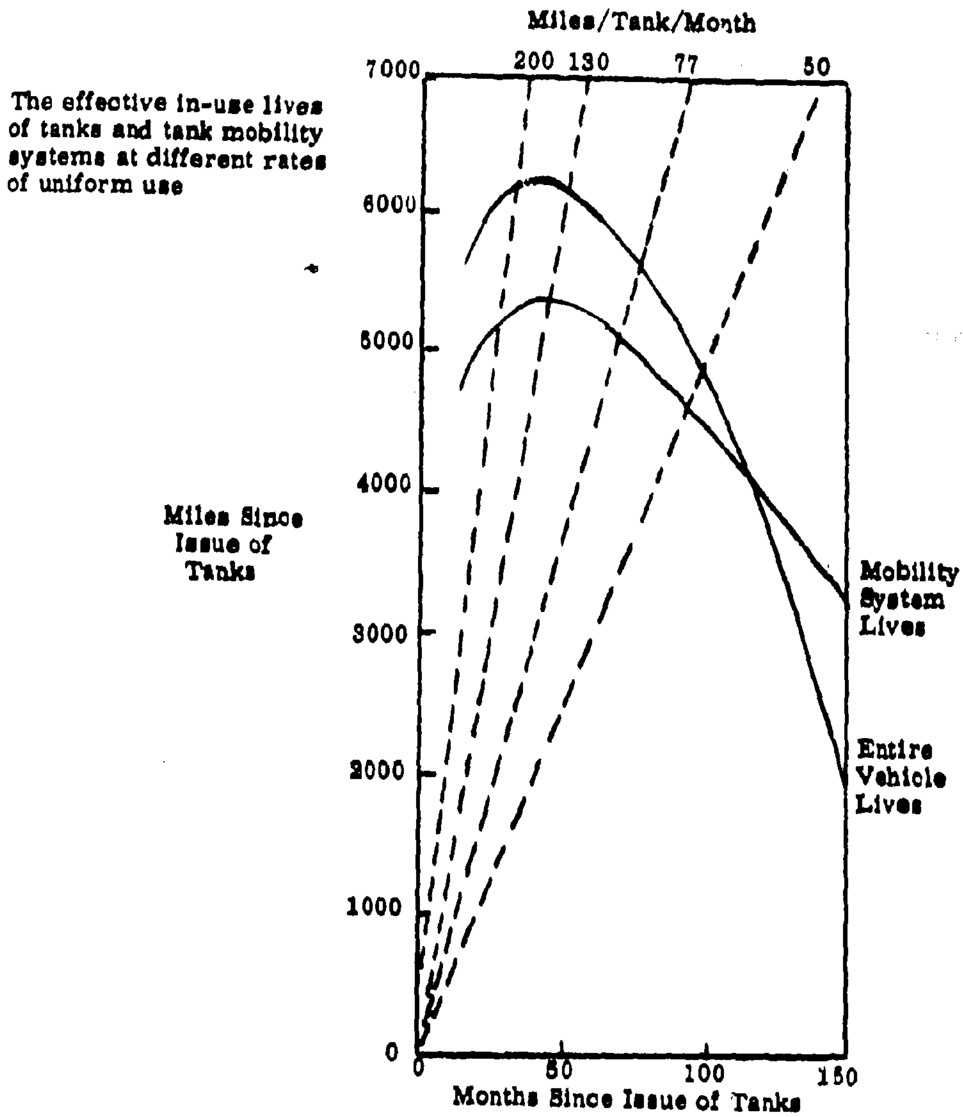
Figure 19

Fig. 20—Money is not the only penalty of vehicular old age. Even though parts are fed into the tank fleet, the net availability of tanks drop. In reality tank users have to pay more and get less as their vehicles accumulate mileage. In this chart both the parts costs and unavailability of tanks like those studied are shown. Out of a force of 1000 tanks, the equivalent of more than an entire battalion are on the average unserviceable and unavailable during the second and third years. Compared with the first year, parts costs have increased about five-fold and unavailability about three-fold be the second and third years.

Fig. 21—Practically all the preceding results may be categorized as by-products of the general analysis leading to the determination of target ages for the effective in-use lives of vehicles. Through suitable weighted

Fig. 20–THE DOUBLE PENALTY–
MONEY and DOWNTIME





Miles/Tank/Month
200
130
77
50
The effective in-use lives of tanks and tank mobility systems at different rates of uniform use
7000
6000
5000
4000
3000
2000
1000
0
Miles Since Issue of Tanks
Mobility System Lives
Entire Vehicle Lives
0
50
100
150
Months Since Issue of Tanks

Fig. 21

combination of the foregoing kinds of information schedules of lives were determined for the tank model studied. These lives are best represented by a curve on a mileage-calendar age plane. In general effective in-use life depends on rate of utilization. In fact the presented curve applies only to an operation spectrum of uniform utilization. The general accumulation of mileage may be represented by a path belonging to a parametered family of usages. Each use family will result in a different life curve. As long as the mileage paths of a given use family do not intersect one another, the time mileage plane consists of points associable with at most a single utilization curve, and a single effective life curve can be constructed uniquely. Existing history data do not provide an adequate basis for the construction of a realistic life model employing intersecting mileage paths of single families and leading to life differences at the points of intersection. Such questions are interesting from the programming and function-theoretic point of view, but they remain well beyond current capacities for empirical, experimental resolution.

Two separate effective in-use life curves are shown in the figure. The one applies to entire tanks and the other to separately defined tank mobility systems. In general higher utilization rates may be expected to increase mileage lives except at very high use. However, the higher mileages are achieved in much shorter times. Much of the management problem arises because the training program results in large numbers of old model tanks with relatively low mileages and in equally large numbers of newer model tanks with much higher mileages. It becomes necessary to establish tank life paths through the inventory in such a way that in the long run the less obsolescent tanks are also the ones with the lower mileages. To have obsolete, low-mileage tanks and modern, over-used ones at the same time achieves nothing more than a use-, readiness-, budget-paradox.

RAC has presented a rapid survey of many of the factors considered in a well-integrated program of assistance to the US Army in the management of its combat vehicles. Time and space do not permit treatment of all factors, nor do they suffice for adequate explanation of the integration to final result.

RAC's activities in this area represent a continuing sequence of alternating empirical and theoretical efforts. Over the years the guiding doctrine has been to provide a steady output of information of general

and specific utility to the Army in the management of its combat vehicles. The information has had to be consistent with the materiel experience of US Army troops in combat units undergoing actual training and yet remaining constantly responsible for preserving a combat ready posture. Applicability to high population fleets and not elegance in an artifically reduced inventory has been a test to be satisfied for all resulting conclusions and recommendations.

APPLICATION OF STATISTICS TO EVALUATE SWIVEL HOOK-TYPE CROSS CHAIN FASTENERS FOR MILITARY APPLICATIONS OF TIRE CHAINS

Otto H. Pfeiffer
U. S. Army Tank-Automotive Center
Warren, Michigan

ABSTRACT. The test was conducted according to a developed experimental design to determine the utility of swivel-hook cross-chain fasteners for military applications of tire chains.

Dual tire chain assemblies (mounted on M35A1 test vehicle) consisting of standard Military-type side and cross chains and three types (standard and two swivels) of cross chain fasteners were subjected to an accelerated wear test of 425 miles on dry concrete road surfaces. The experimental results were expressed in terms of: (1) Miles to failure for an individual cross chain, (2) Weight losses of selected cross chains and (3) Replacement times for each of three types of fasteners. The principal response, (1) miles to failure, was considered to be exponentially distributed; therefore, logarithms of miles to failure were analyzed in accordance with the structure of the experimental design.

Cross chains connected with swivel-type fasteners remained functionable about twice as long as the cross chains connected with the standard-type fasteners. Both swivel-type fasteners permitted significantly faster cross-chain replacement than the standard type, although one swivel type was also significantly faster to manipulate than the other swivel type.

The statistical analysis of the experimental results indicated the swivel hook-type cross chain fasteners used in this test resulted in a significant increase of cross chain life as well as simplification of replacement.

INTRODUCTION. It became necessary for the government to make a decision whether to consider swivel hook-type cross chain fasteners in future procurement of tire chains and components. Some information on swivel hooks was available, but it was considered inadequate for the basis of a decision.

A test was proposed to utilize two military trucks equipped with various types of tire chains, to be conducted on both concrete and gravel road surfaces (in a two-to-one proportion) for a total distance of approximately 300 miles.

Products of two manufacturers of tire chains and chain components were available for the test. Each had a swivel hook of a particular design which they were interested in selling to the Government.

Since a large proportion of the Military's wheeled cargo vehicles fall within the $2\frac{1}{2}$ ton payload class (see Figure 1) having 9.00 x 20 size tires, it was desirable to test tire chains of this dimension.

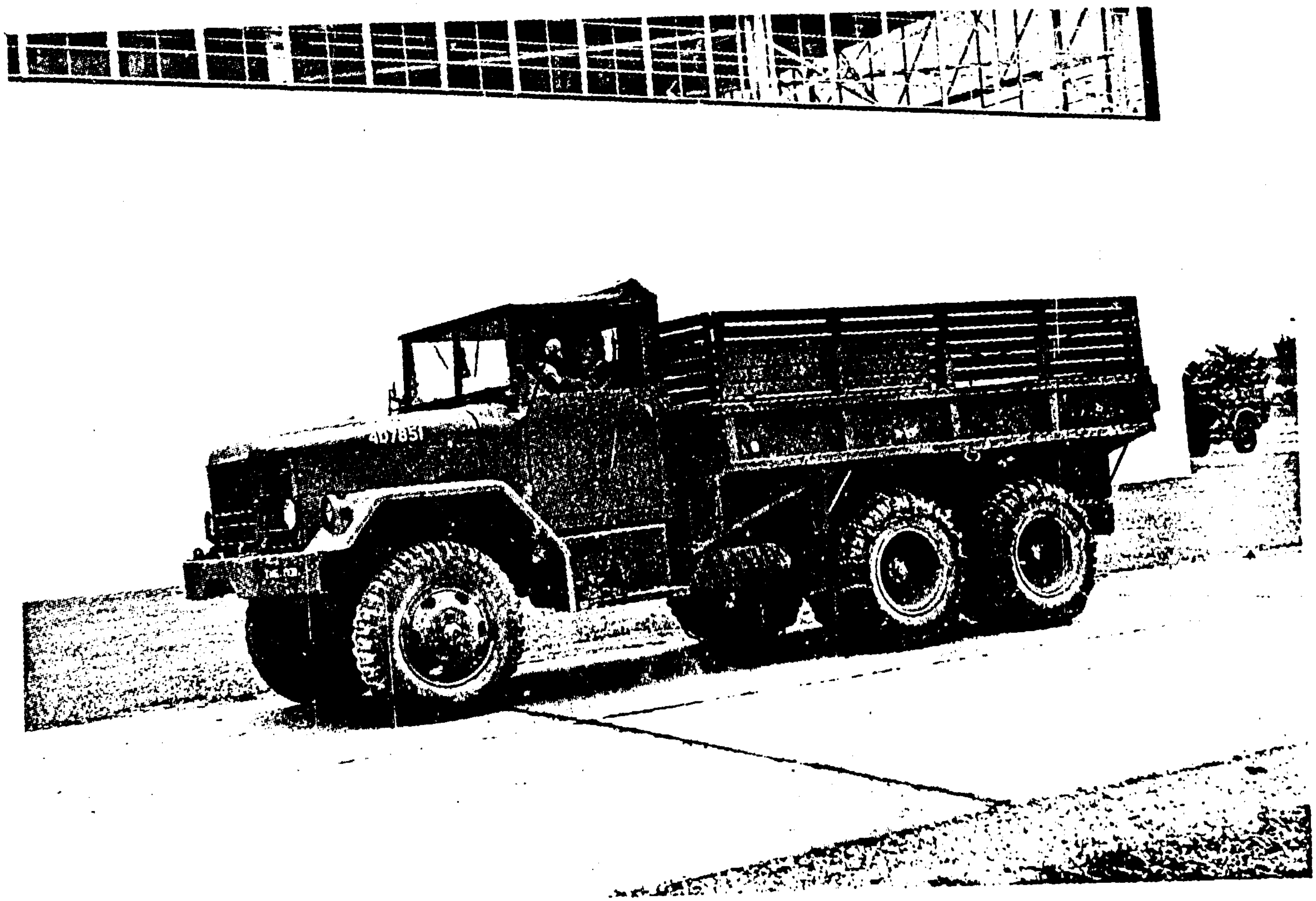
It was requested that representative cross chain samples of all test chain assemblies from the two manufacturers hereafter referred to as "Code B" and "Code C" and a standard Military item referred to as "Code A" be subjected to a metallurgical examination to determine: mechanical properties, macro-etch quality, case-hardened depth, and cross-sectional hardness.

Two complete tire chain assemblies of Code B were to be placed on two rear dual wheels of one vehicle, diagonally opposite to two Code A chain assemblies. The same arrangement utilizing Code C and Code A was to be adhered to on a second test vehicle.

During the course of testing, it was proposed that two brake panic-stops per mile be made after the vehicle had attained maximum speed. Wheel spining on take-off was also requested.

A review disclosed a test [3] had previously been conducted on tire chains by the Government. This test compared the Code B and Code A chain only. A similar chain arrangement to this proposed test was used, but the vehicle was driven 500 miles over various terrains and on surfaces ranging from marsh and swamp to concrete. Although this test indicated some advantages of the swivel hooks in comparison with the standard Military chain, the test did not adequately establish the quantitative nature of the advantages.

Only letters of recommendation from commercial sources were available regarding the Code C chain.



U. S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO. 72275 DATE 11 June 63
Tire Chain Test. M35A1, 2 1/2 Ton Truck, Cargo, 6x6. Left Front View of Truck with Chains Installed.

FIGURE 1

The similarity between the newly proposed test and the test previously conducted brought up the question: What could be learned with this test that had not been found out before? The only answer that appeared obvious was -- Nothing! Further study of the situation anticipated many problems if the proposed test arrangement and procedure were to be followed:

1. Differences in the vehicles, not only in weight but manufacturing variances as tire sizes, tracking characteristics, and braking characteristics, to mention a few.

2. Differences in the tire chains, both length of cross chains as well as the cross-chain wire diameter and of major concern -- the difference in metallurgy of the cross chain steel.

3. The difference in road contact surfaces when chains are compared on different wheels, or worse -- on different vehicles.

4. Difficulty of data analysis or supportable conclusions if the test were to be conducted in a haphazard manner or without accounting for known variables.

We decided a specially-designed plan must be developed which would produce usable data. The services of Dr. Emil H. Jebe, a Research Mathematician from the University of Michigan's Institute of Science and Technology, were obtained. He became an inseparable part of the project until final conclusions were reached.

<u>CHOICE OF RESPONSE AND EXPERIMENTAL UNIT</u>. It was apparent that several responses should be studied. Of primary interest were:

1. Miles to failure of each individual cross chain,

2. The weight loss of the cross chains associated with each type of cross chain fastener,

3. The time needed for replacement of worn-out or broken cross chains.

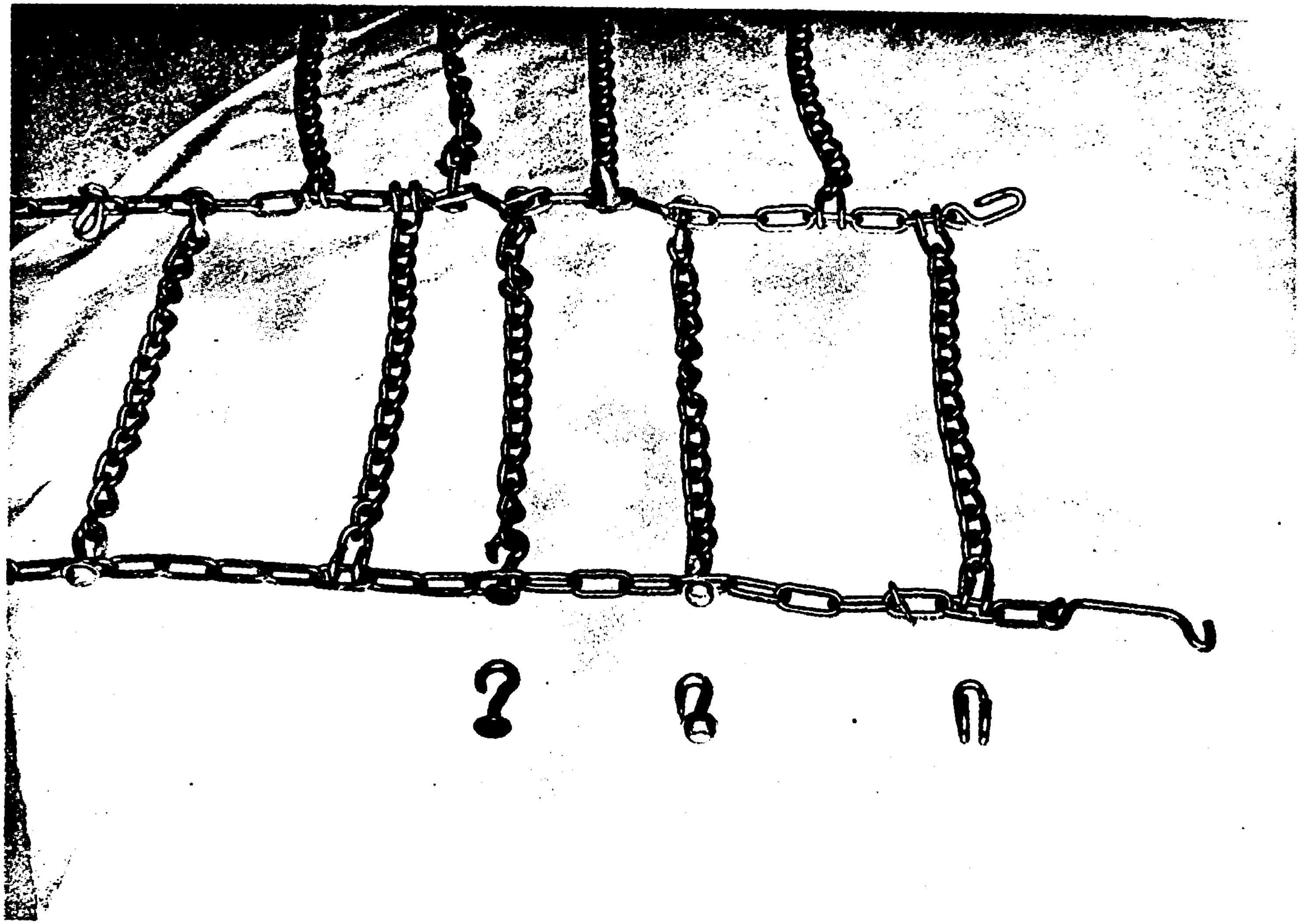
The first major problem in development of a suitable experimental design of plan was to determine the unit for measuring responses. The four rear dual wheels and the inside and outside tire of each dual wheel were the first kinds of units considered. However, since a single tire chain covers the complete dual wheel (resulting in only four units being available at the same time), no satisfactory design could be based on a wheel as the unit unless 1200 to 2000 miles could be driven. Even using a single tire as the unit would have given only eight units and differences between outside and inside tires would have to be eliminated. This kind of unit would also require an extended period of driving.

Further discussions disclosed that the government was considering at this time only the utilization of swivel hooks as repair items for the ample supply of standard Code A chains presently in stock.

An interesting thought occurred -- why not use standard Code A tire chains with Code B and Code C swivel-hook fasteners inserted as if they were repair items? This would eliminate several of the anticipated problems. The cross chains of Code A or standard military tire chains were considered to be for all practical purposes of uniform size and metallurgical composition.

It suddenly became evident that with this concept, a large number of experimental units would become available if we considered each individual cross chain as the unit of measurement. A dual tire chain consists of 26 cross chains or 13 on each half. A total of 104 units became available which could be used for one run of, say -- 300 to 500 miles.

Some minor problems were encountered in the acceptance of this experimental unit. The swivel hook-type fastener could not be inserted into the same link of the side chain as the standard crimp hook. It required an adjacent link 90° out of phase. This problem was solved by spacing the cross chains unevenly around the tire as shown in Figure 2. Since three types of cross chain fasteners were to be tested, four cross chains were fastened with each type. The remaining cross chain was left fastened with the standard type and was not used for test data. Using this arrangement, there were 32 cross chains for each fastener-type, evenly distributed over the eight tires of the four dual wheels.



U.S. ARMY TANK-AUTOMOTIVE COMMAND NEG. NO. 71825 DATE 1 Apr 63
Modified standard military dual tire chain utilizing Code A, Code B, and Code C cross chain fasteners

FIGURE 2

A completely randomized arrangement of the cross chain fasteners on each wheel seemed undesirable. A randomized starting order followed by a systematic order was suggested. The cluster of four cross chains connected with the same type fastener on each tire became the experimental unit with eight replications over the set of wheels.

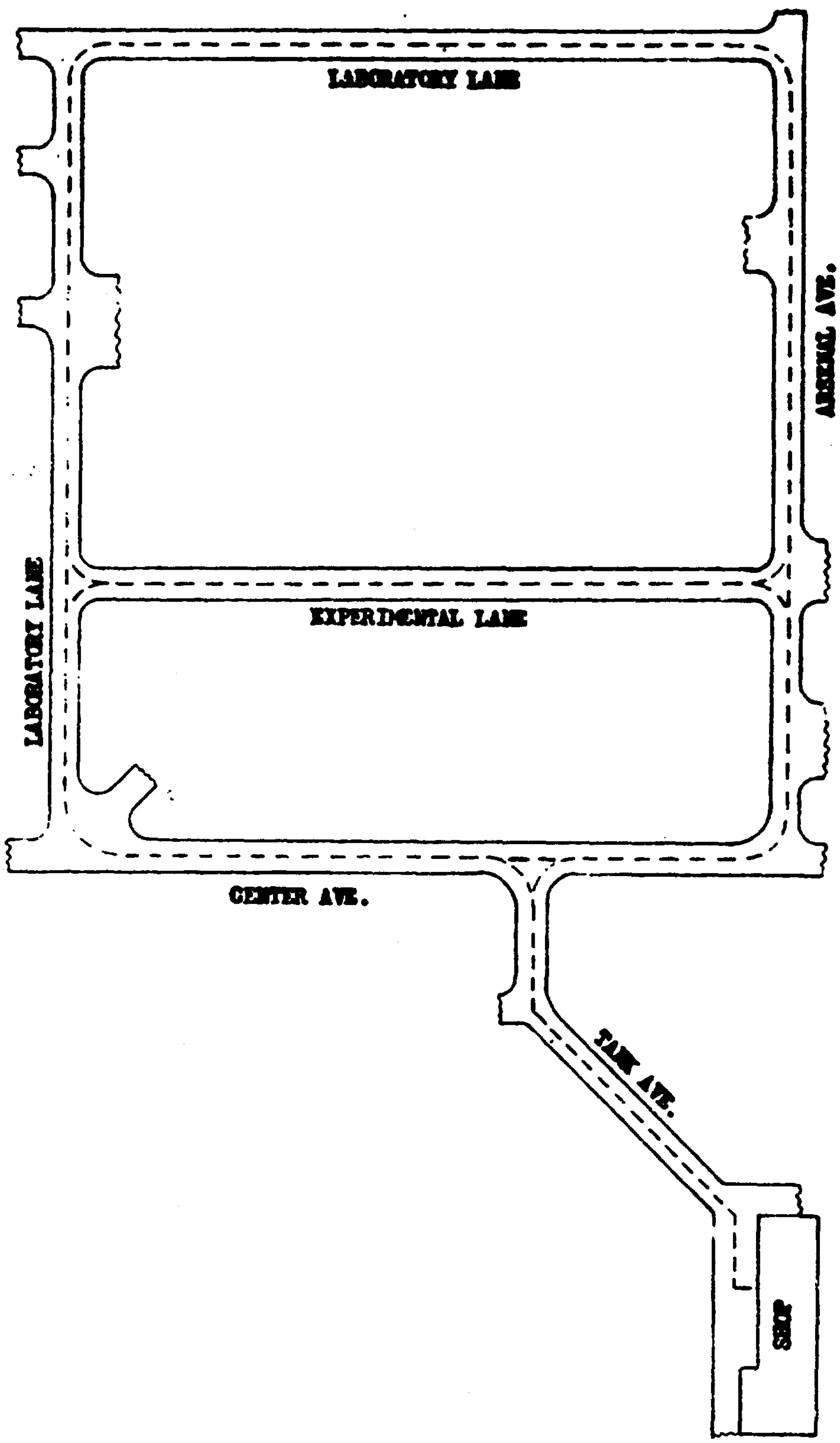
Four chain assemblies were fabricated according to prescribed procedure and mounted on the test vehicle. To accelerate the rate of wear and thereby reducing driving distance, a test course (see Figure 3) consisting entirely of concrete was selected. Provisions were made to equalize right and left turns necessary during test driving.

STRUCTURE OF THE DESIGN. With the experimental unit now clearly defined, it was possible to develop the complete overall plan for data analysis. There were certain obvious sources of environmental variation present which could not only be removed but estimated for magnitude. These sources are listed as:

1. Difference between front and rear dual wheels.
2. Difference between right and left side dual wheels.
3. Outside versus inside tires.
4. Interactions of these effects with each other.
5. Possible interactions of the treatments (types of cross chain fasteners) with these positional differences.

Since there were three clusters - three experimental units - of four chains (each cluster with different types of fasteners) on each tire the plan may be described as a Randomized Complete Block Design in eight replicates considering each tire as a block. The variation among blocks was also to be subdivided in the manner just outlined.

A formal structure for the design could then be established. The usual textbook model for a Randomized Complete Design would be satisfactory for preparing an analysis of variance of these observed results, providing the usual assumptions could be made. One of these



TIRE CHAIN WEAR TEST COURSE

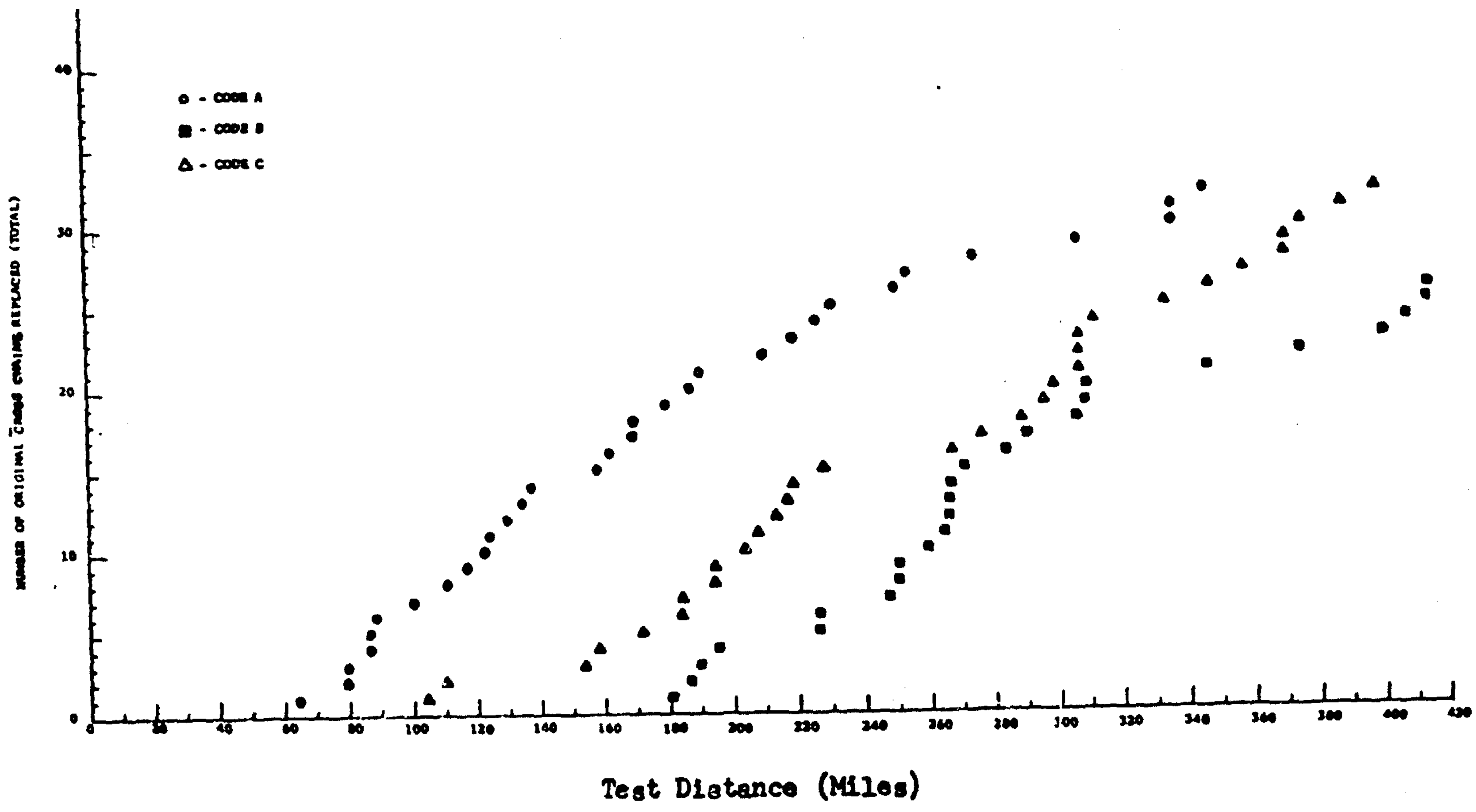
FIGURE 3

assumptions is that the observations are independently and normally distributed [1] [10]. The response of primary concern here is miles to failure of an individual cross chain. Therefore, the test plan may well be considered a "life test" or a "wear test". Considering that our observations were "miles to failure", we realized they would not likely be well-described by the normal probability distribution. This point is well demonstrated in Figure 4. It appears therefore that the exponential distribution may be regarded as an acceptable probability model for the observed miles to failure. With this in mind, our analyses of these failure data have been generally guided by the considerations set forth in a series of papers by B. Epstein, B. Epstein and M. Sobel, and M. Zelen which appeared in several Mathematical and Statistic Journals (See reference list).

The exponential distribution in its probability density form usually expresses the random variable in some quantity directly related to time. In our case, d = miles-to-failure may be used as the random variable. In this form, a constant uniform failure rate is assumed. As was indicated previously, we do not have a uniform situation in this tire chain test since there are a number of sources of variation present. The parameter θ appearing in the probability density form represents the mean time before failures of the cross chains. Based on Zelen's work [19], a more complex model (Figure 5) was written for this θ in the exponential distribution.

Another view may be taken for observations following exponential distribution. The procedure established here suggests taking logarithms of the observations, in this case miles-to-failure for each cross chain, and then carrying out an analysis of variance considering necessary assumptions. The functional relation between mean and variance for exponential distribution, that is, the standard deviation equals the mean [2] suggests the use of the logarithmic transformation (see Figure 6). This approach is also discussed by Zelen in his paper [19]. He finds the technique acceptable against possible departures from the strict exponential distribution form.

The analysis of variance of the logarithms was prepared on the basis of the model and will be discussed later.



MILEAGE AT FAILURE OF ORIGINAL CROSS CHAINS

FIGURE 4

EXPONENTIAL DISTRIBUTION MODEL

The model $\theta_{ijkr} = \theta\alpha_i \gamma_j \tau_k \rho_r \eta_{ij}$ was developed from works by M. Zelen. This model states the Mean Time Before Failure (MTBF) for the four crossbars secured with the same type of fastener on the same tire depends upon

θ	a general mean time to failure
α_i	an effect associated with the front or rear set of dual wheels, i=1, for Front, and i=2, for Rear
γ_j	an effect associated with left or right side of truck, i.e., j=1 for Left, j=2 for Right
τ_k	an effect associated with the tire position, k=1 for Outside, and k=2 for Inside
ρ_r	an effect associated with the type of cross-chain fastener, say 0_A, 0_B and 0_C, or the subscript r takes three values
η_{ij}	an interaction effect associated with the dual wheel position

FIGURE 5

LOGARITHMIC TRANSFORMATION MODEL FROM EXPONENTIAL DISTRIBUTION FORM

Let $Y = \log d$ where d is miles to failure.

The transform model is written as:

$$Y_{ijkrs} = m + a_i + c_j + (ac)_{ij} + t_{ijk} + f_r + e_{ijkrs}$$

where

m	is an overall mean logarithm
a_i	is the front or rear effect
c_j	is the left or right side effect
$(ac)_{ij}$	is an interaction or component associated with i, jth wheel position
t_{ijk}	is the tire position effect within the i, jth wheel
e_{ijkrs}	is a random deviation for the sth individual cross-bar within the rth type and its associated tire and wheel position

FIGURE 6

The test was originally planned to run for 300 to 500 miles. At 300 miles, the test was temporarily stopped to assess the situation up to that time. It was already clear that there were large differences in miles to failure for types B and C versus type A even though approximately 1/3 of the original cross chains had not yet failed. This large number of unfailed chains would have created considerable difficulty when analyzing the results. Driving was continued to about 425 miles before being terminated for other reasons. At that time, six of the original cross chains remained. These were all equipped with type B fasteners and were positioned as follows: One on each of two wheels, and four on one wheel, as shown in Table I. It was necessary to estimate data values for these six in order to maintain a balanced and simple, straight-forward analysis of the results. In general, procedures derived from work by B. Epstein [8] were used for estimating the missing data. Using Epstein's formula (see Figure 7) we obtained estimates for the missing values on the two single tires. In the first case, $n = 4$, $r = 3$, $d_r = 309.8$ and $d_1 = 226.6$. Solving for $\hat{\theta}$ we obtained the value of 123.25. This formula assumes that failures occur randomly at any time starting from zero miles of travel. Since failures did not occur for some distance of travel, we estimated the minimum "guarantee" distance A by the formula $\hat{A} = d_1 - \hat{\theta}/_n$. In this case, $\hat{A} = 195.79$. Combining $\hat{A}$ and $\hat{\theta}$, we obtained 319.04 as the proper estimated MTBF* for the cell. Using the estimated cell mean, we estimated the cell total as n (estimated MTBF) = 4(319.04) = 1276.16. We already knew the actual miles to failure of three of the cross chains in the cell; therefore, subtracting this value from the estimated total left 433.1 as the estimated missing value.

The same method was used to estimate the second missing value at 504.3.

No failures of B type swivel hook fasteners were evidenced on the outside tire of the left front wheel. This situation posed a real problem. The Epstein formula used for estimating the previous two missing values requires at least two failures in a cell if it were to be applied directly. A variety of methods for solving this problem were considered, including schemes based on using the weights of the unfailed cross chains at termination of the test driving. All these schemes were rejected as unsuitable.

*Mean Time Before Failure

	LEFT			RIGHT		
	A	B	C	A	B	C
	MILES	MILES	MILES	MILES	MILES	MILES
Front Outside	336.4	-	289.2	336.4	414.6	312.4
	275.0	-	153.9	180.1	400.8	347.7
	250.6	-	371.3	86.8	414.8	194.7
	346.8	-	334.3	64.7	346.8	208.3
Front Inside	254.1	309.8	158.3	219.3	266.1	269.7
	169.9	-	184.1	122.4	271.9	307.7
	158.3	306.7	358.4	129.8	247.1	399.6
	88.5	226.6	307.7	110.6	186.5	171.8
Rear Outside	231.5	250.6	183.9	307.7	266.8	110.1
	86.8	195.6	213.9	162.6	308.6	376.1
	134.5	264.8	388.8	116.9	-	204.1
	187.5	309.2	217.2	79.6	375.1	228.6
Rear Inside	211.5	259.8	371.0	190.8	250.6	296.7
	169.7	180.6	104.3	79.6	267.0	194.4
	100.0	189.9	276.9	124.0	291.0	307.7
	137.7	284.6	299.6	226.3	226.5	219.0

MILES TO FAILURE OF ORIGINAL CROSS CHAINS

TABLE I

EPSTEIN'S FORMULA

$$\hat{\theta} = \frac{\sum d_i + (u-r)d_r - nd_1}{r-1}$$

where (in this particular case)

$\hat{\theta}$ = estimated MTBF

d_i = observed miles to failure for ith failed item

n = number of items under test

r = number of failed items

d_r = the largest value or last failure

d_1 = the smallest value or first failure

FIGURE 7

In order to obtain a useable solution, the Epstein formula was reapplied to the whole of Type B fastener data. Utilizing the two previously estimated missing values, the parameters now became n = 32 and r = 28 in this instance. The mathematics will not be described here. But stating briefly, $\hat{\theta}$ and $\hat{A}$ were estimated, then $32(\hat{A} + \hat{\theta}) - \sum_{i=1}^{28} d_i$ yielded an estimated total for the entire missing cell. This estimated total /4 provided the estimated MTBF or $\hat{\theta}$ for the cell. Individual values were then determined by proportionality of each cross chain weight to mean weight of the cell. The estimated missing values for this cell ranged between 563 and 572 miles. These values appeared to be too high, giving the impression we were favoring type B. Applying the standard analysis of variance "missing plot procedure" for the Randomized Complete Block Design to the logarithms of the miles to failure data [14] [18] provided a mean value for the entire cell although it was not a useable value. The estimated value based on averaging the data available was 276 miles, but it was known these cross chains had already traveled over 400 miles. The distance traveled at termination of test, 424.9 miles, could also have been assigned to each unfailed cross chain but this would have been unfavorable to type B. Therefore, values estimated as already described were used and they appeared to yield an acceptable solution.

There are several approaches which may be followed in considering the estimation of the effects of interest in this experiment. For completeness, three methods were considered and the differences among the methods were small for this test program. The methods considered were:

1. Calculating the appropriate simple averages of the miles to failure data

2. Estimating the parameters in Zelen's model as described above

3. Estimating in terms of averages of the logarithms of miles to failure.

The latter of the three methods was used in the analysis of variance and will be discussed further.

Logarithms of the original data, miles to failure, and the anti-logs of the mean logarithms are shown in Table II.

Averages of Logarithms of Miles to Failure
and the Anti-logs of these Averages

Effect	Tire Position					
	Inside		Outside		Combined	Average
	Log Miles	Miles	Log Miles	Miles	Log Miles	Miles
Fastener Type:						
A	2.1682	147.3	2.2381	173.0	2.2032	159.7
B	2.4088	256.3	2.5744	375.3	2.4916	310.2
C	2.3974	249.7	2.3871	243.8	2.3923	246.8
Wheel Position:						
Left Front	2.3549	226.4	2.5554	359.3	2.4551	285.1
Left Rear	2.3004	199.7	2.3191	208.5	2.3098	204.1
Right Front	2.3220	209.9	2.3795	239.6	2.3508	224.3
Right Rear	2.3220	209.9	2.3455	221.6	2.3337	215.6
Sides:						
Left	2.3276	212.6	2.4372	273.7	2.3824	241.2
Right	2.3220	209.9	2.3625	230.4	2.3423	219.9
Wheels:						
Front	2.3385	218.0	2.4673	293.3	2.4029	252.9
Rear	2.3112	204.7	2.3323	214.9	2.3218	209.8
Overall Mean:						
	2.3248	211.2	2.3999	251.1	2.3623	230.3

TABLE II

Comparing the average for type A fastener (inside tires only, expressed in anti-log form as 147.3 miles with values calculated by the other two methods -- 155.8 miles and 155.9 miles, respectively) we find it slightly less. This somewhat lower figure is the result of the non-linearity of the logarithmic transformation. This anti-log value is really an estimator of a median value rather than a mean.

Estimating the differences among the types of cross chain fasteners, or the ratios as, say -- B/A, C/A and B/C was the next concern. It is not a simple problem and considerable study was devoted to finding a reasonable solution. A method discussed by Zelen [19] [20] for estimating such ratios and finding confidence limits for the ratio gave extremely wide limits for these ratios from our experimental data. From the averages of "log miles" for "Insides Tires", "Outside Tires" and "Combined Averages" presented in a previous table, the fastener-type differences expressed in logarithms were calculated. The variances of these differences were estimated directly by taking 2(Experimental Error Mean Square) /r, where r is the number of values averaged to form a mean for a single fastener type. This was based on the theorem that the variance of a difference is the sum of the variances of the quantities used to form the difference. The standard deviation was obtained from the variance result just stated. Confidence intervals were then formed by taking the observed mean differences: $(B-A) \pm t_{\alpha k}$ (standard error) where t in this case was $t_{(0.95,86)} = 1.987$. The k = 86 degrees of freedom comes from the Pooled Error Mean Square determined from the analysis of variance presented later. The confidence intervals obtained are for differences of averages expressed in logarithms.

There was a considerable difference in the average miles to failure (about 50 miles) considering all the fasteners combined between the Inside Tires and the Outside Tires. Considering this difference, it was decided to present separate results for Inside and Outside Tires and then combined averages, as seen in Tables II and III. Returning results to original scale of miles to failure was desirable. It was observed that the anti-log of the difference (B - A on Inside Tires) was nearly the ratio of miles to failure for B/A as given by the data in Table II. Anti-logs taken for the lower and upper confidence limits for the differences likewise became approximate confidence limits for the ratios. These results are also listed in Table III.

Estimated Type Differences in Logarithms and Estimated Ratios of Miles to Failure by Types of Crossbar Fasteners and the Associated Confidence Intervals

Description Difference	Ratio	Estimate	Lower Limit	Upper Limit
Inside Tires				
B-A		0.2406	0.1289	0.3523
	B/A	1.7400	1.3460	2.2510
C-A		0.2292	0.1175	0.3409
	C/A	1.6950	1.3110	2.1923
B-C		0.0114	-0.1003	0.1231
	B/C	1.0260	0.7938	1.3280
Outside Tires				
B-A		0.3363	0.2246	0.4480
	B/A	2.1690	1.6770	2.8050
C-A		0.1490	0.0373	0.2607
	C/A	1.4090	1.0900	1.8230
B-C		0.1873	0.0756	0.2990
	B/C	1.5390	1.1900	1.9910
Combined				
B-A		0.2884	0.2095	0.3673
	B/A	1.9430	1.6200	2.3300
C-A		0.1891	0.1102	0.2680
	C/A	1.5460	1.2890	1.8540
B-C		0.0993	0.0204	0.1782
	B/C	1.2560	1.0480	1.5070

TABLE III

It is to be noted that these confidence limits, when expressed in the original scale in miles, are really confidence limits for a median mileage ratio figure.

ANALYSIS OF TOTAL VARIATION. The analysis of variance in terms of logarithms of miles to failure is presented in Table IV.

The selected chain arrangement as previously described was a cluster of four cross chains for each type of fastener systematically arranged around each tire. When comparing error, it was found the error mean square (based on the clusters) was about equal to the mean square for the cross chains within the clusters (except on the inside tires where there was some difference, but in the wrong direction). It was decided to calculate the pooled error.

The results in Table IV bear out the large differences between the types of fasteners already displayed in Tables II and III. Other points to be noted from Table IV are:

1. Left side versus right side effect is small (about equal to error).

2. The front wheels versus rear wheels effect is large in relation to error although this effect is mostly associated with the outside tires.

3. Individual wheels differ considerably from what might be expected if predictions were based only on the left versus right and front versus rear effects. This effect is shown by the interaction line which is again largest for the outside tires.

4. The difference in miles to failure for outside tires versus inside tires, about 40 miles, does not appear to be a chance effect. Most of this difference was associated with the left front wheel, however.

5. During the detailed examination there was some question regarding the uniformity or behavior of the types of fasteners on the outside tires versus inside tires. Although not shown in Table IV, the effect of an interaction of types by outside versus inside was calculated (removed

Analysis of Variance of Logarithms of Miles to Failure of Cross Chains in Test of Cross Chain Fasteners for Truck Tire Chains

Source of Variation	Degrees of Freedom	Inside Tires	Outside Tires		Combined Analysis
		Sums of Squares			
Total	48	260.8730	278.9757	(96)*	539.8537
Mean	1	259.4277	276.4531		535.7455
		Mean Squares			
Outside vs Inside				(1)	0.1353
Front vs Rear	1	0.00897	0.21902		0.1583
Left vs Right	1	0.00373	0.06701		0.0388
Interaction	1	0.00564	0.12268		0.0987
Wheels				(3)	(0.0986)
Tires				(7)	(0.0803)
Fastener Types	2	0.29490	0.45442		0.6872
Error	6	0.00701	0.03021	(14)	0.0248
Crossbars within Clusters	36	0.02222	0.02844	(72)	0.0253
Standard Error of a Difference	=	.0562	.0562		.0397

$$\text{Pooled Error} = \frac{14(.0248) + 72(.0253)}{86} = .025249$$

Standard Deviation = 0.1589

*Quantities in ( ) are degrees of freedom for the combined analysis.

TABLE IV

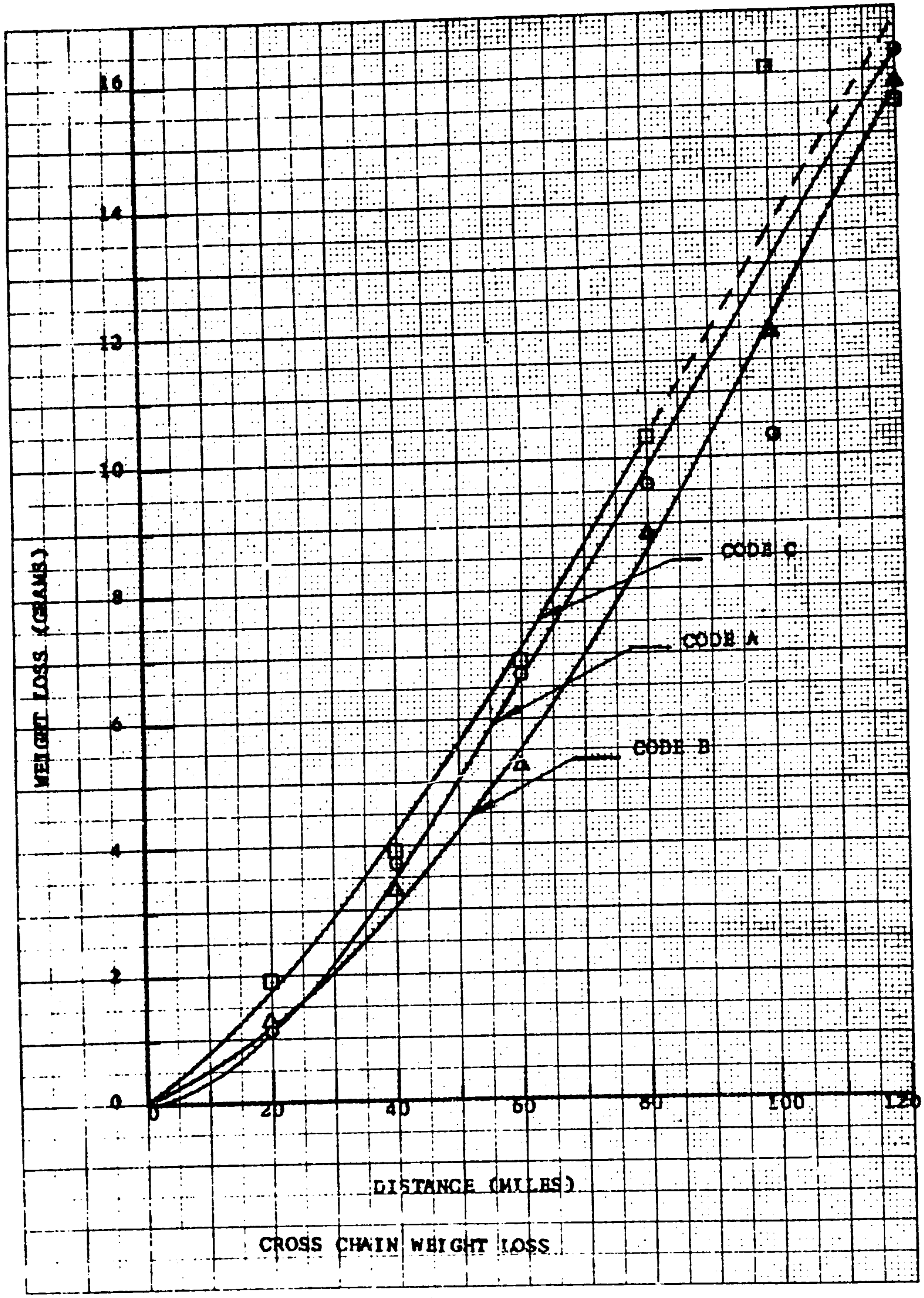
from error with 14 degrees of freedom). An F ratio of about 2.46 was obtained when comparing the mean square obtained with pooled error mean square. The probability of such a value of more extreme occurring by chance was about .09.

OTHER TESTS. During the temporary termination of the test at 300 miles, a Median Test described in works by Mood [13] was applied to the data. From this test at the 300 mile level, it was reasonable to conclude that B and C type fasteners were much superior to type A; however, it seemed desirable to perform additional test driving to further quantify the experimental results.

OTHER RESULTS OF INTEREST.

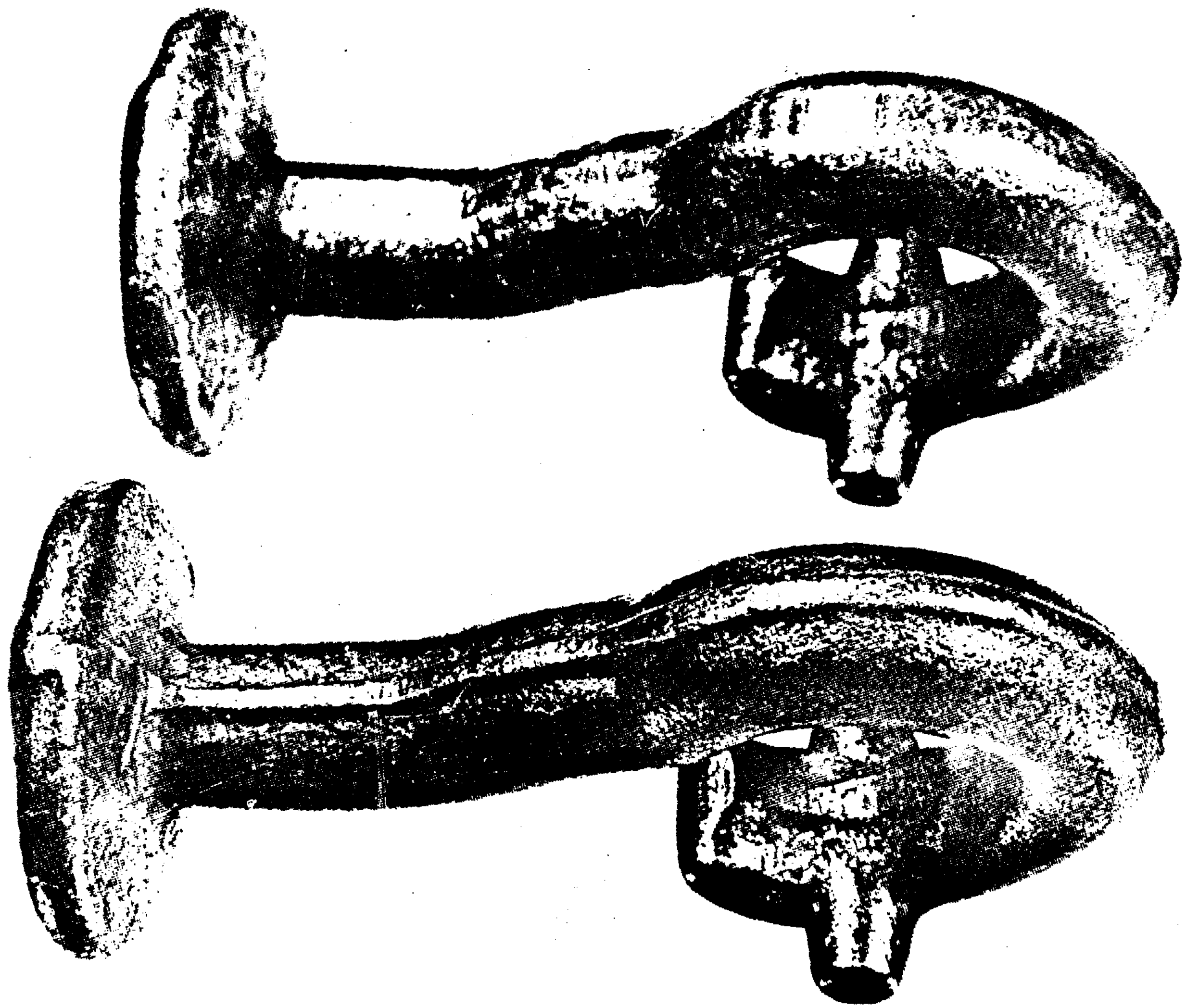
1. Weight loss:

Weight loss measurements for individual cross chains were made at 20-mile intervals during the test program while original chains remained intact. To remove correlation between successive weighings, different cross chains were selected for weighing at the end of each 20-mile interval. At each weighing, one cross chain for each type of fastener was removed from each wheel and then replaced in its original position. With the limited number of cross chains available, it was necessary to repeat weighting at the end of 180 miles. The resulting weight loss data was plotted against distance driven (see Figure 8). These weight loss data show that all cross chains (regardless of fastener type) tended to lose weight at approximately the same rate. Considering this result, the longer life of cross chains associated with the B and C type fasteners must be due to a spreading of the wear over the entire surface of the cross chain produced by the rotational motion permitted by the swivel-type fasteners. Many of the type C swivel fasteners had a forge flashing on the hook shank (shown in Figure 9) which restricted the rotational motion. Depending on whether there were two, one, or none of these hooks with "flash", a particular cross chain might rotate freely, only partially (wind-up), or not at all. Such results could account for the observed difference in life of the B and C types. Figures 10, 11, and 12 show specific wear patterns associated with each type of fastener. The curved wear pattern established in the C type is assumed to be the result of one non-rotating hook causing chain wind-up.



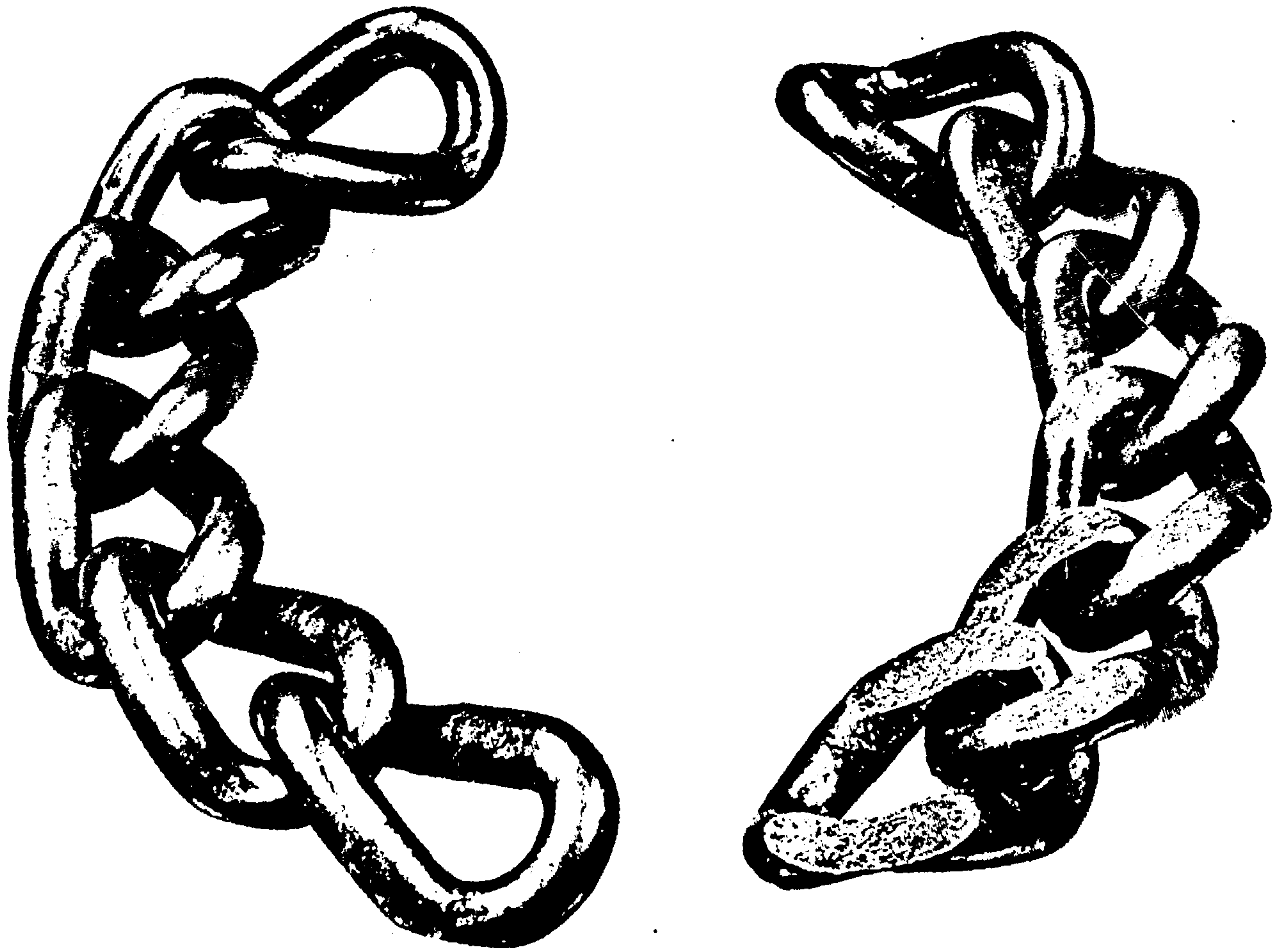
WEIGHT LOSS (GRAMS)
16
14
12
10
8
6
4
2
0
0
20
40
60
80
100
120
CODE C
CODE A
CODE B
DISTANCE (MILES)
CROSS CHAIN WEIGHT LOSS

FIGURE 8

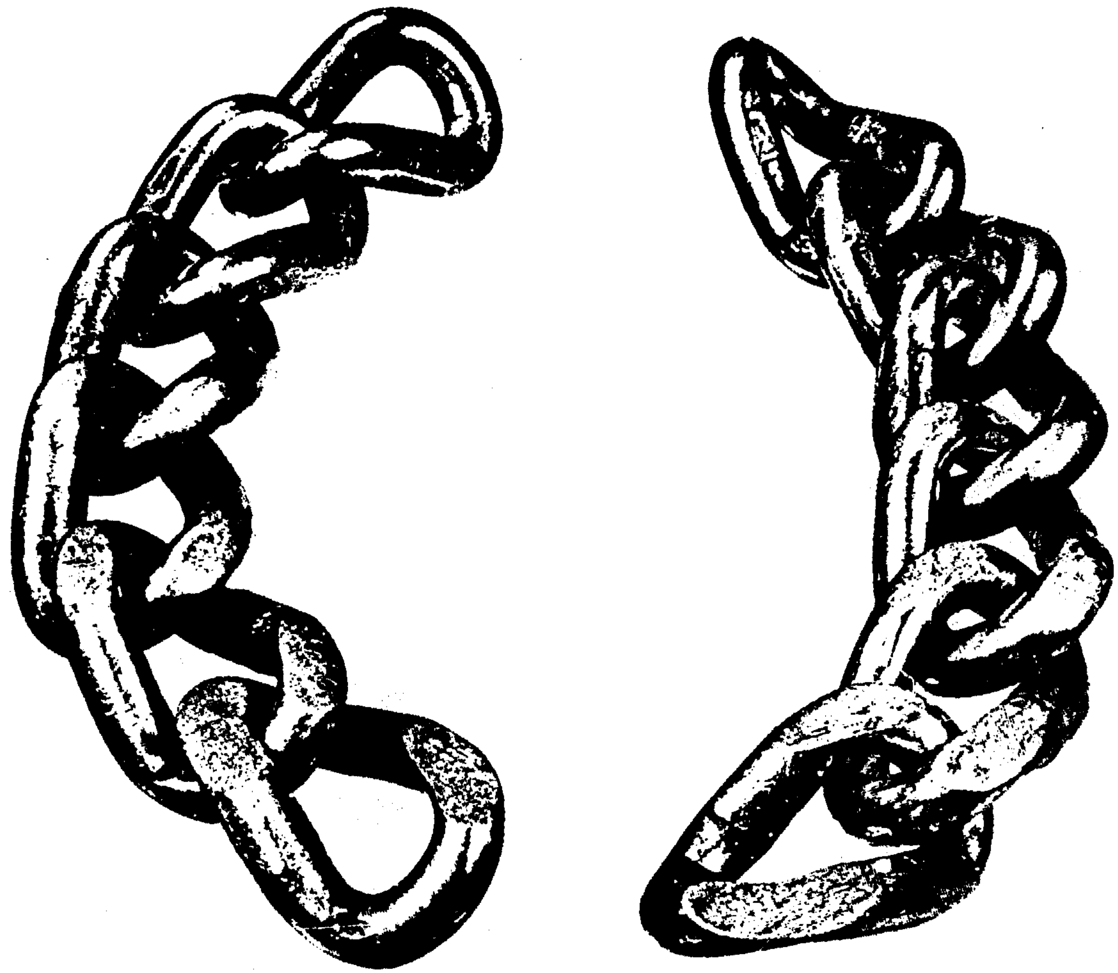


U. S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO. 72213 DATE 20 May 63
Tire Chain Wear Tests. Non-uniformity of Code C connectors. "Flash" present on hook.

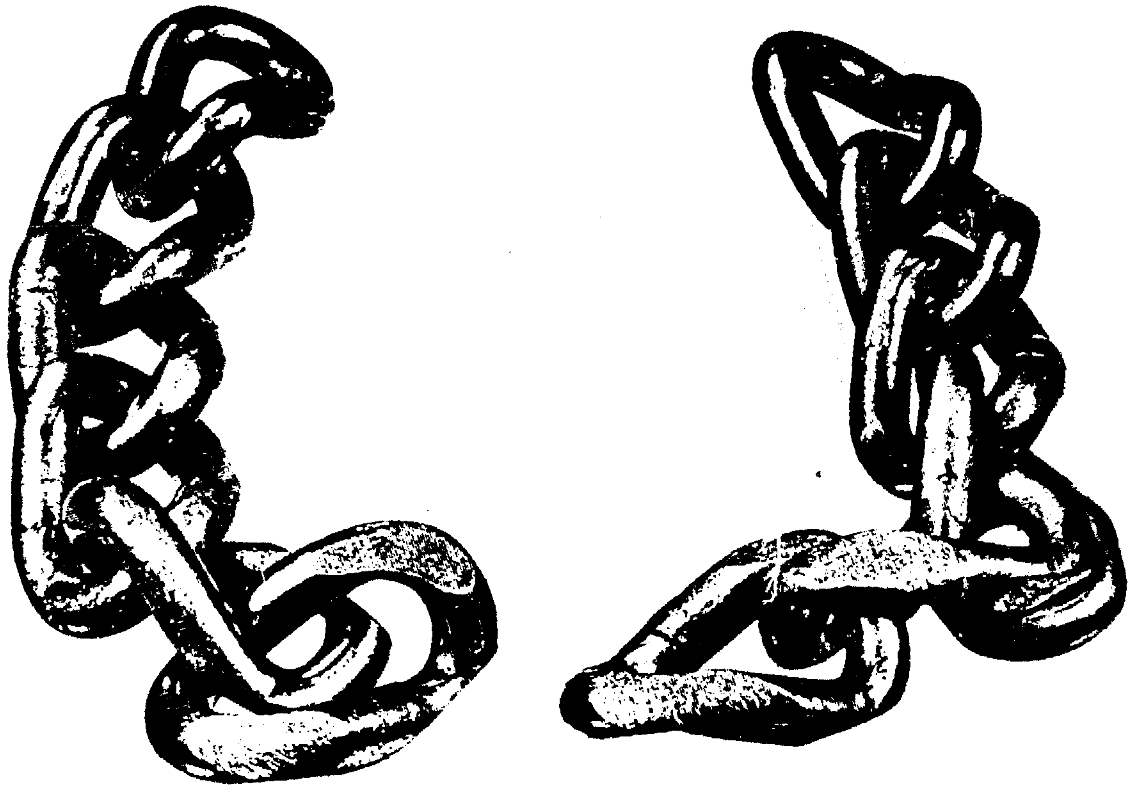
FIGURE 9



U.S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO. 72215 DATE 20 May 63
Tire Chain Wear Tests. Typical wear at failure of cross chain connected with Code A type fastener. One-half of chain shown upside down.
FIGURE 10



U. S. ARMY TANK-AUTOMOTIVE CENTER **NEG. NO.** 72216 **DATE** 20 May 63
Tire Chain Wear Tests. Typical wear at failure of cross chain connected with Code B type fastener. One-half of chain shown upside down.
FIGURE 11



U. S. ARMY TANK-AUTOMOTIVE CENTER **NEG. NO.** 72218 **DATE** 20 May 63
Tire Chain Wear Tests. Typical wear at failure of cross chain connected with Code C type fastener. One-half of chain shown upside down.

FIGURE 12

2. Replacement Time:

At each replacement of a failed cross chain and also during the weighing procedure in the shop, the time required for removal and reinstallation of the cross chain was recorded. A separate record was kept for shopwork and field work. There was some difference associated with the work site; however, the most pronounced difference was associated with type of fastener. An analysis of variance of these differences could have been calculated, but the mean differences in observed time between the three types were so large that a detailed analysis did not seem to be needed. The data were analyzed using the Wilcoxon-Mann-Whitney Statistic (ranking method) [12] [15]. There was found to be a significant difference between the averages in all cases. The average replacement time and re-installation time and ratios are tabulated in Table V.

All removals and installations of the A-type fastener were done with a special tool (see Figure 13) provided for this test. An unsuccessful attempt was made to remove an A-type fastener with the tools provided in the standard tool kit of the vehicle.

CONCLUSIONS. It is clear from the results obtained that the cross chain fasteners type B and C (swivel hook) are superior to the standard type fastener. This superiority is primarily described by comparing average miles to failure or by the ratios of average miles to failure. From these ratios for inside tires only, we observe the swivel-hook fasteners to be about 70 percent better on the average. On the outside tires, we find type B about 117 percent better than type A, and type C 55 percent better than type A. Confidence limits for the true ratios of superiority show a minimum of at least 30 percent improvement on inside tires, and possibly as much as 120 percent for the swivel hooks when expressed in terms of two-sided 95 percent confidence intervals. These results were far less uniform on the outside tires.

The next question raised is, "Are the type B swivel hooks better than the type C?" The observed difference on the inside tires is small. On the outside tires the data indicate a significant difference between type B and C. A large part of the superiority of B and C is found on the left front outside tire. Type B is also better than C on the other three outside tires, but to a varying degree.

REPLACEMENT TIME AVERAGES AND RATIOS

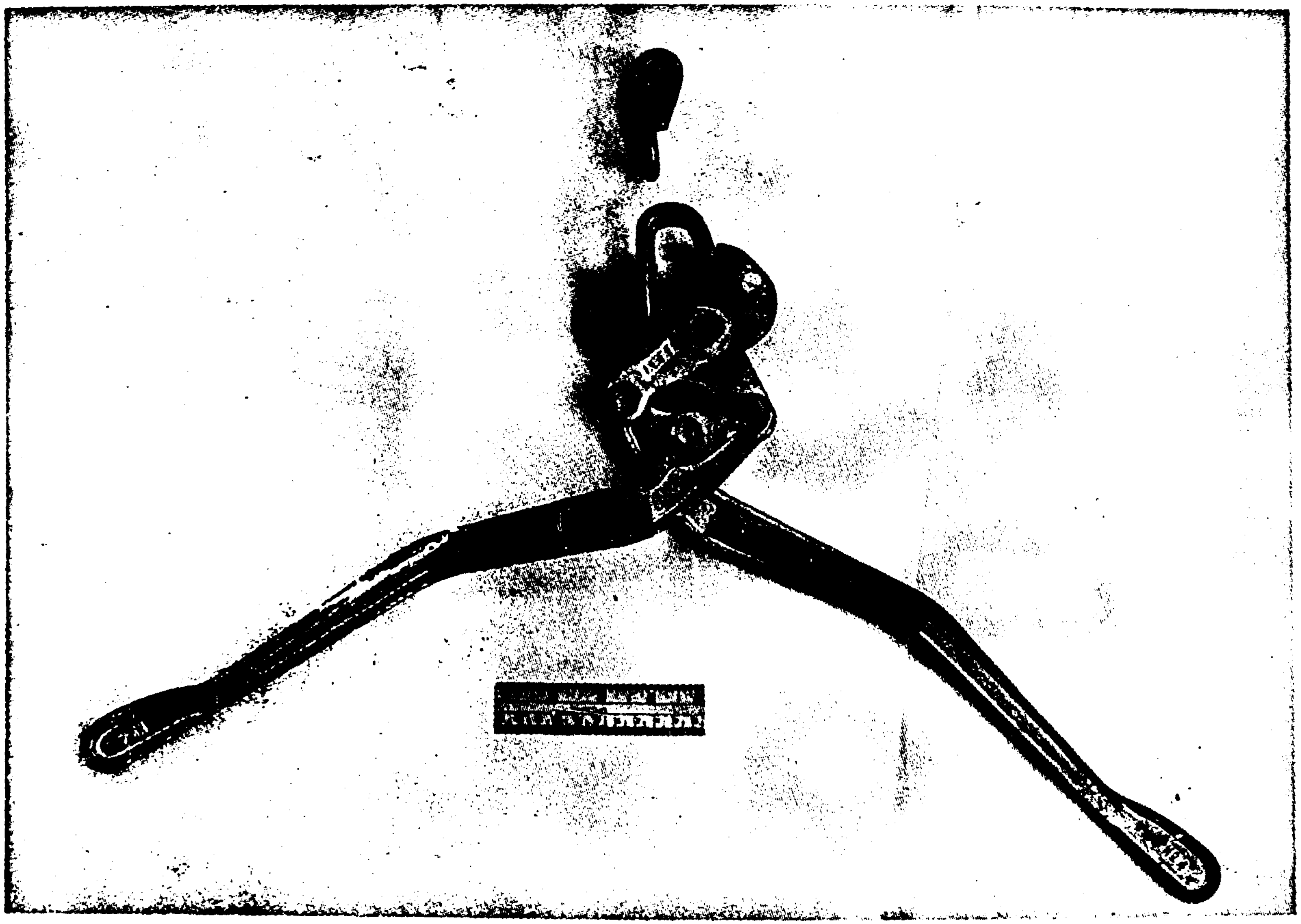
REPLACEMENT OF FAILED CHAINS IN THE FIELD

	CODE A	CODE B	CODE C
Average Time in Seconds	261.4	83.1	55.2
Ratios	A/B= 3.15	A/C= 4.74	B/C= 1.51

REMOVAL AND RE-INSTALLATION OF CROSS CHAINS IN SHOP

	CODE A	CODE B	CODE C
Average Time in Seconds	225.5	98.0	40.5
Ratios	A/B= 2.30	A/C=5.57	B/C= 2.42

TABLE V



U. S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO.P/R 243-64-3 DATE 2 August 63
Special Chain Tool with Code A Fastener in "Ready to Close" position.
FIGURE 13

As described before, it was noted the type B fastener allowed better cross-chain rotation than the type C so that part of the difference between B and C may be ascribed to this characteristic, although the sizeable difference in performance for outside and inside tires is baffling.

The replacement times are highly favorable to the swivel hooks although type C was found to be somewhat better than B. Thus, it seems that minor modifications might make the two swivel hooks about equal in performance and replacement time.

A field trial conducted using chains completely assembled with swivel hooks would be worthwhile to determine the extrapolation factor for normal field conditions from the accelerated test conditions.

When considering the use of the swivel-hook type fasteners as replacements in military tire chains, it appears from the data obtained that the present experiment has been adequate.

REFERENCES

AMS = Annals of Mathematical Statistics
ASTM = American Society for Testing Materials
JASA = Journal of the American Statistical Association

1. Anderson, R. L., and Bancroft, T. A., "Statistical Theory in Research," McGraw-Hill, New York, 1952.

2. Bartlett, M. S., "The Use of Transformations," Biometrics, 3, 39 (1947).

3. Cuevas, R. N., Jr., "Test of Removable Swivel Hooks for Tire Chains," Report No. DPS/TT1-649/53, August 1957, Dev. and Proof Services, APG, Md.

4. Davies, O. L., "Design and Analysis of Industrial Experiments," Oliver and Boyd, Edinburgh, 1954.

5. Epstein, B., "Truncated Life Tests in the Exponential Case," AMS. 25, 555 (1954).

6. Epstein, B., and Sobel, M., "Life Testing I," JASA, 48, 486 (1953).

7. Epstein, B. and Sobel, M., "Some Theorems Relevant to Life Testing from an Exponential Distribution," AMS, 25 373 (1954).

8. Epstein, B., "Estimation of the Parameters of Two Parameter Exponential Distributions from Censored Samples," Technometrics, 2, 403 (1960).

9. Epstein, B., "Estimation from Life Test Data," Technometrics, 2, 447 (1960).

10. Kempthorne, O., "The Design and Analysis of Experiments," J. Wiley and Sons, New York, 1952.

11. Laurent, A. G., "The Lognormal Distribution and the Translation Method . . . ," JASA, 58, 231 (1963).

12. Mann, H. B. and Whitney, D. R., "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," AMS, 18, 50 (1947).

13. Mood, A. M., "Introduction to the Theory of Statistics," McGraw-Hill, New York, 1950.

14. Neyman, J. and Scott, E., "Correction for Bias Introduced by a Transformation of Variables," AMS 31, 643 (1960).

15. ORDNANCE Corps Pamphlet 20-113, Ordnance Engineering Design Handbook, Experimental Statistics, Section 4, Special Topics, Ordnance Corps, June 1962.

16. ORDNANCE Corps Pamphlet 20-114, Ordnance Engineering Design Handbook, Experimental Statistics, Section 5, Tables, Ordnance Corps, June 1962.

17. Parker, W. H., "The Wearing Qualities of Tire Chains," ASTM Procedures 28 (II), 332 (1928).

18. Snedecor, G. W., "Statistical Methods," 5th Edition, Iowa State College Press, Ames (1956).

19. Zelen, M., "Factorial Experiments in Life Testing," Technometrics, 1, 269 (1959).

20. Zelen, M., "Analysis of Two-Factor Classifications with Respect to Life Tests," paper No. 42 in "Contributions to Probability and Statistics, Essays in Honor of H. Hotelling," edited by I. Olkin, Stanford University Press, 1960.

ERROR ANALYSIS PROBLEMS IN THE ESTIMATION OF SPECTRA

Virginia B. Tipton
White Sands Missile Range, New Mexico

ABSTRACT. Power spectral density functions are estimated digitally by evaluating the Fourier cosine transform of the autocorrelation function. In order to obtain reliable averages with which to describe the autocorrelation function it is necessary to limit the resolution with which it, and its transform, can be described. Is it possible to evaluate, or to express analytically, the accuracy with which the computed spectrum represents the true spectral density function?

INTRODUCTION. The use of power spectral density functions to describe the frequency content of a time function has been a common engineering practice for some time, developing originally from the communications engineers' concern with separating signals from noise in transmission systems. At the same time the statisticians' approach to the study of random fluctuations in time series data led to the development of autocorrelation functions as a descriptive tool. The bridge between these two approaches to the study of noise, which is simply high frequency random variations superimposed on the desired data, was the discovery of the now well-known Wiener-Khinchin relationship. This relationship simply states that, except for a constant factor, the power spectral density function and the autocorrelation function of a stationary random process are a Fourier transform pair. Since the autocorrelation function is an even function of its time lag τ, the complex Fourier transformation process simplifies to a real cosine transformation which can easily be carried out by a digital computer.

The digital computation of power spectral density functions is becoming an increasingly more important part of data reduction work. It is now being applied experimentally to the study of random errors in trajectory measuring instrumentation systems, as well as to the more traditional applications in vibration data analysis and telemetry problems.

However, in order that the spectral estimates computed may be of value to the data user, we must be able to describe in some way the reliability with which the computer spectrum approximates the true spectral density function; that is, we must be able with some degree of confidence to place limits upon the errors of our estimation.

THE WSMR DATA REDUCTION SPECTRUM ANALYSIS PROGRAM. The derivation of the computer programming equations used at WSMR Data Reduction Directorate was first given in a report written in 1957, "The Digital Computation of Power Spectra," by L. M. Spetner of Johns Hopkins University. This digital process is based on the Wiener-Khinchin relationship; that is, it first computes the autocorrelation function of the random data and then determines its Fourier transform, which is the power spectral density function. In order to separate the noise data from any constant (zero frequency) component, the input data are first averaged and then this data mean is subtracted from the original data. This process insures that the average of the residuals will be zero, a condition which must be met if the Fourier transform is to exist. In order to eliminate any linear trend, or a quadratic, a least squares 2nd degree curve is then fit and removed. We are now ready to compute the autocorrelation function of the residuals, which we assume then to be both random and stationary.

At this point it is well to say a few words about random processes in general. A random process is a collection, or ensemble, of time functions such that the ensemble can be characterized by its statistical properties. In studying noise problems we are usually not overly concerned with that individual time function which we happened to observe, since any of the member functions of the ensemble could have occurred with equal probability. Rather we are interested in determining from the observed function the statistical properties which characterize the entire ensemble. For a special class of random processes (that is, for those which are both stationary and ergodic) this can be done because it has been shown (elsewhere) that in such cases the process averages across the ensemble are equal to the time averages along a single representative function from the ensemble (See Figure 1).

The autocorrelation function for a random process is defined as the ensemble average of the product of each function times itself shifted by a time delay τ.

$$R(\tau) = \widetilde{f(t) \cdot f(t+\tau)}, \tag{1}$$

where the wavy bar indicates averaging across the ensemble.

If we are dealing with a single random function from an ergodic ensemble, equation (1) for the autocorrelation function becomes a time average over the function

$$(2) \qquad R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(t)\, f(t+\tau)\, dt$$

In the digital case where the integral is replaced by a summation over the range of data points N and time delay m, this becomes

$$(3) \qquad R(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} f(i)\, f(i+m) \; .$$

This autocorrelation function has several interesting properties:

(1) It is an even function, i. e., $R(-\tau) = R(\tau)$ (a property which is useful in determining its Fourier transform.)

(2) The value of $R(\tau)$ for $\tau = 0$ equals the average power of $f(t)$, or in statistician's language, the variance of the function.

(3) The value of $R(\tau)$ is bounded by its value at $\tau = 0$, so that the computed autocorrelation coefficients can easily be normalized to give unity autocorrelation for zero time delay.

If the function is truly random then its autocorrelation function will rapidly approach zero, since the values of $f(t+\tau)$, as τ increases, are not dependent upon the value of $f(t)$. Thus a typical normalized autocorrelation curve of a random noise record will have the shape indicated in Figure 2.

However, it should also be pointed out that the converse is not so - it cannot be shown that because the autocorrelation function approaches zero as τ increases that the given function is necessarily random.

Once the autocorrelation function has been found, the power spectral density function is computed from it by taking its Fourier transform

(4) $$\phi(\omega) = \int_{\tau=-\infty}^{\infty} R(\tau) e^{-i\omega\tau} \, d\tau \, .$$

Using the property that $R(-\tau) = R(\tau)$, this becomes

(5) $$\phi(\omega) = 2 \int_{0}^{\infty} R(\tau) \cos \omega\tau \, d\tau \, .$$

The spectrum is estimated for discrete values of $\omega = \frac{\pi K}{m}$, where K is an index ranging from 0 to m, and m the number of autocorrelation coefficients computed.

The resulting estimates are smoothed using a 3-point symmetric filter, with weights (0.23, 0.54, 0.23) and plotted as function of frequency.

(For reference, the digital computing formulas used in the program will be listed as an appendix to the paper.)

THE PROBLEM OF ERROR ANALYSIS. The problem confronting us now is chiefly this: How can we express the errors involved in estimating spectra by this digital process? Or in other words, with what confidence can we say that the spectrum we have computed represents the true spectrum of the process we are studying? Can we put limits on our error, perhaps in the form of a statement such as, "our estimate is within $\pm 5\%$ of true spectrum" and have perhaps 90 or 95% confidence that we are right?

The problem appears to be in balancing the frequency resolution we can achieve, that is, the number of points used to estimate the spectral curve, against the reliability with which they are computed. The maximum frequency resolution (Δf) in our digital process is determined by the highest frequency we can distinguish in the data (f_{max}) and the number of time delay averages (M) for which we computed the autocorrelation function.

$$\Delta f = \frac{2 \cdot f_{max}}{M}$$

But the highest distinguishable frequency is limited by the rate at which the original data samples were digitized. The sampling rate as given in Spetner's equations must be at least twice the highest frequency present.

By the time the data arrive in digital form at the Data Reduction Computer facility we no longer have any control over the digitizing rate $(1/\Delta t)$ or the length of the data sample [T(seconds) = (N points) $\cdot$ (Δt seconds)]. We must assume that the data users chose a sampling rate high enough to minimize aliasing errors, that is, the folding back of frequencies higher than f_{max} so that they appear as some sub-multiple of themselves in the frequency range we can observe.

In addition, the number of time delay averages used to describe the autocorrelation function is limited by the length of the data sample. In practice, we generally limit M to approximately one-tenth of the number of data points N. $(M \doteq N/10.)$ We could increase the number of time delay averages computed, but only at the cost of reliability of them. As M increases the number of data points available to average decreases. Thus this could not solve our problem, and at present, no other solution has been found.

DIGITAL COMPUTING FORMULAS

1. Autocorrelation function is estimated at M points

$$R(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} f(i) \cdot f(i+m), \text{ for } 0 \le m \le M.$$

2. Cosine transforms estimated for each value of M

$$L(0) = R(0) + 2 \sum_{P=1}^{M-1} R(P) + R(M)$$

$$L(h) = 2R(0) + 4 \sum_{P=1}^{M-1} R(P) \cos \frac{h P \pi}{M} + 2 R(M) \cos (h\pi),$$

$$\text{for } 0 < h < M$$

$$L(M) = R(0) + 2 \sum_{P=1}^{M-1} (-1)^P R(P) + (-1)^M R(M).$$

3. Smoothed spectral estimates

$$u(0) = 0.54\, L(0) + 0.46\, L(1)$$

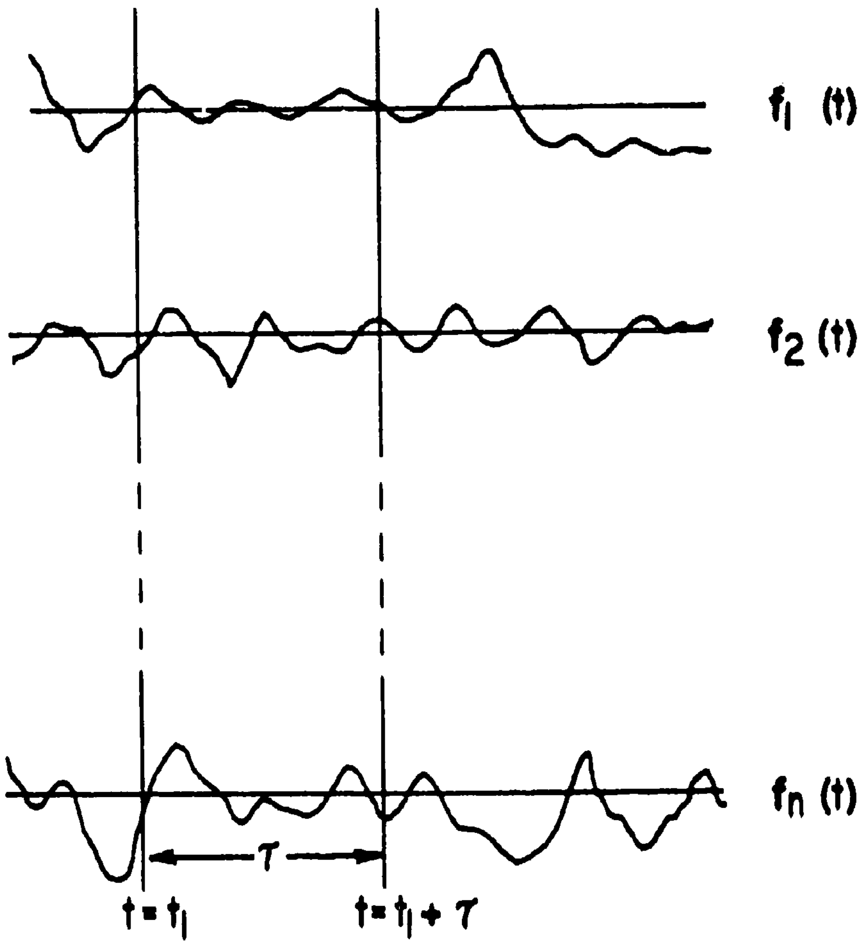
$$u(h) = 0.54\, L(h) + 0.23\, [L(h-1) + L(h+1)] \text{ for } 0 < h < M$$

$$u(M) = 0.54\, L(M) + 0.46\, L(M-1)\ .$$

4. Plot of $u(h)$, for $0 \le h \le M$, where Δt is the time interval in seconds between original data points, vs, frequency, $f(h)$, where

$$f(h) = \frac{h}{2m\, \Delta t} \text{ cycles per second.}$$

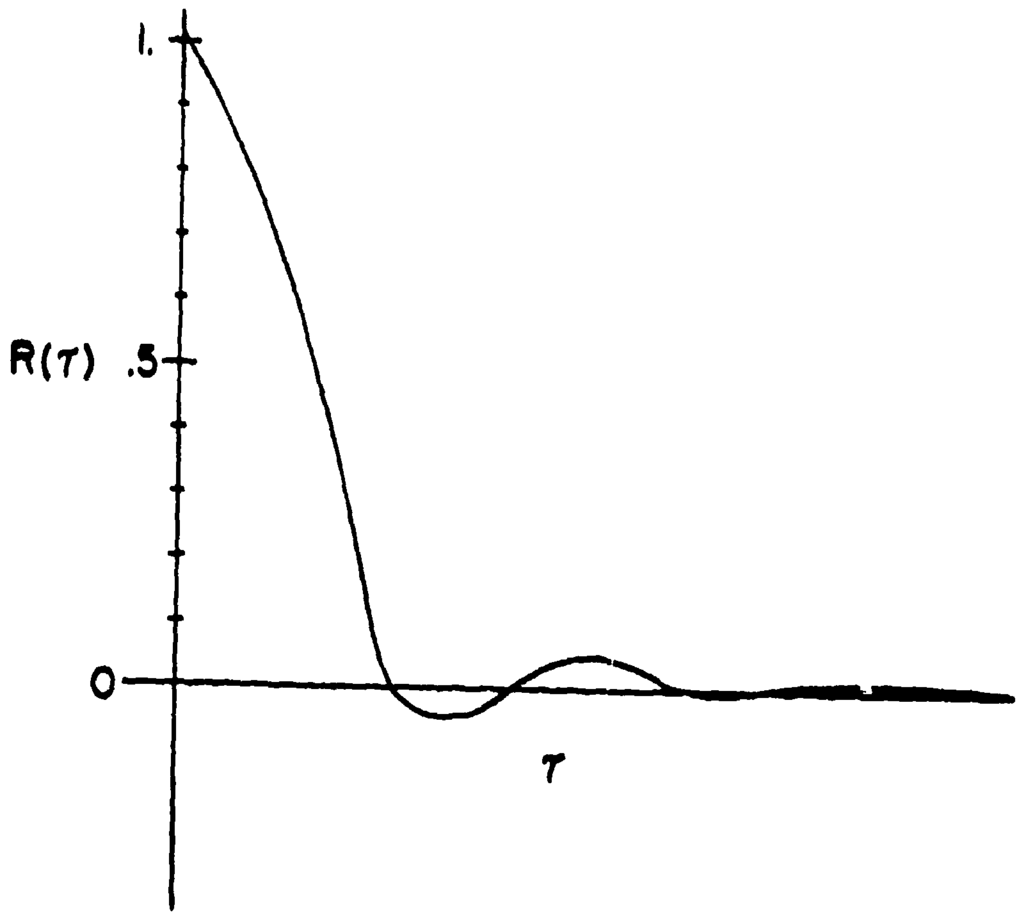
AN ENSEMBLE OF n TIME FUNCTIONS



ENSEMBLE AVERAGE $\quad \widetilde{f(t=t_1)} = \frac{1}{N}\left[f_1(t_1) + f_2(t_1) + \cdots + f_n(t_1)\right]$

TIME AVERAGE $\quad \overline{f_2(t)} = \frac{1}{N(\Delta t)}\left[f_2(t_1) + f_2(t_2) + \cdots + f_2(t_n)\right]$

FIGURE 1.



A TYPICAL NORMALIZED AUTOCORRELATION FUNCTION FOR RANDOM NOISE

FIGURE 2.

VALIDATION PROBLEMS OF AN INTERFERENCE PREDICTION MODEL

William B. McIntosh
U. S. Army Electronics Proving Ground

The Electromagnetic Environmental Test Facility (EMETF) of the US Army Electronic Proving Ground, Fort Huachuca, Arizona, is being developed to give solutions to a host of communications-electronic problems, most of which in one way or another arise from the fact that military demands imposed upon the electromagnetic spectrum require vastly more space in the spectrum than is available for this purpose. A consequence of the resulting crowding is interference between electromagnetic equipments. The EMETF is designed to provide experimental data bearing on the interference problem in its broadest sense. A subsidiary and included feature is provision of data on the ability of communications-electronic systems to perform intended missions in the absence of competing electromagnetic signals.

Although the ultimate EMETF will be a facility capable of providing these answers for communications equipment, for radar, for navigation devices, for data transmission links, and in fact, for all army electronic activities, this presentation will be confined to aspects of voice communication by radio.

Several years ago, the EMETF was conceived of as primarily a huge outdoor field test facility, spread over some 2400 square miles. In the District of Columbia area, this facility would have stretched from one side of Washington to the other side of Baltimore. An artist's concept of the field facility is shown in figure 1. Originally, 24 transmitter sites and two transmitter-receiver siteswere deployed; the latter two sites were the test sites. At each of the transmitter sites equipment was grouped around a control van. The master control center is located at one of the test sites. From the center, one or more transmitters at any one or more van sites may be controlled for test purpose.

The basic test unit is a cycle, figure 2, which requires 30 seconds; it consists of energizing the desired transmitter, and recording the test link performance, as will be described later. Then the entire environment is turned on, and the link performance again measured. If degradation has occurred, the next step - and a long one - consists of a search for the one or more transmitters responsible for the degradation. These operations produce the basic field data.

The concept of test, and with it the field facility, has evolved to the point where now the field facility has been reduced in size, and the major share of useful output is being derived from a computer simulation program known as the Interference Prediction Model (IPM). The IPM requires a considerable amount of input data of various sorts, which has lead to the creation of a third unit, the Instrumentation Workshop.

The ultimate form of the IPM, as it is currently envisioned, is shown in figure 3.

The model requires input data specific to the equipment or concept under test as well as a specific description of the problem. Internally, the model consists basically of five modules. Figure 3 also shows in block diagram form how each module is developed and validated.

The propagation module performs one basic function. It describes the attenuation which is expected as an electromagnetic wave front travels between a transmitter and a receiver.

The equipment module incorporates the necessary equations which describe what happens to electrical signals as they pass through the equipment. The scoring module translates the processed signal at the receiver output into a measure of how much of the original intelligence remains. The tactical deployments module incorporates several preselected deployments, each containing the physical location, in three dimensions, of every piece of emitting or receiving equipment, plus information on the topography in the form of an XYZ matrix, in meters, at 500 meter intervals in the XY plane. The radio frequency assignment also becomes a part of this module.

The statistical module is at present largely undeveloped. In time, however, it will be used to convert an essentially deterministic model into a stochastic one. The essential purpose of this paper is to describe certain problems which have been encountered in an attempt to provide an interim stochastic capability by different methods.

I will next outline a test problem in which the important outputs will be obtained from the IPM, but for which actual hardware of the test item is available for certain measurements.

The preparation consists of establishing the location of all equipments, both of the test type and other types which will share the environment, the assignment of radio frequencies, the inputs of the equipment characteristics, and the like. From the entire collection of communication nets in the chosen deployment, a sort of stratified random sample is chosen for test. Stratification is based upon the frequency of different type nets as well as on the relative importance of net type to mission success. Thus, if there is only one command net from Corps to Division, that net is included. From the many command nets between, say, infantry platoons and squads, several are chosen at random.

In the present use of the model, one basic question is asked. What, on some scale, is the overall system effectiveness? This question is often asked for a standard system and for a proposed replacement, whereupon the comparison will provide useful information to those who make procurement decisions. The systems effectiveness measure now in use depends upon a somewhat involved procedure which results in every test link being classified as providing or not providing acceptable performance. Then the effectiveness measure is merely an index, being the ratio of acceptable links to the total tested. Since the initial measure, intelligibility, is changed from a continuous variate to a binomial, many links which have been measured inaccurately will still be classified correctly. Further, to the extent that the model is imprecise but lacking in bias, errors of classification in one direction will tend to be balanced by other errors in the opposite direction. Thus, for the index, we really do not need to be concerned in great detail with the goodness of our answer for each link.

But people do ask such questions. And ultimately we would like to answer such queries as: How well can some specific platoon communicate with its company headquarters? We no longer will be satisfied with knowing how well on the average a platoon can communicate with a company, nor will we accept a simple yes - no answer.

Given this ultimate desire to answer questions about any communications link, of necessity we must accept a stochastic answer. Even if we have developed a perfect model, in the technical sense, this will be true. Communications equipment will continue to exhibit interunit variability; operators will not all have identical hearing ability or training; and most of all, propagation loss will continue to be an important variate. Even if we should come to know the form and moments of all pertinent distributions of equipment characteristics, we won't have any way of knowing

the individual characteristics of the specific equipment at any given geographic point. Anyway, in reality, those equipments may be a few to a few hundred meters at least away from where we have them located in the problem. Even though we learn all there is to know about the effects of atmospheric conditions, terrain, and vegetation on path loss, we can't know what the precise atmospheric factors would be if the tactical operation we are simulating were to exist. Nor can we precisely describe the minutiae of the terrain over all direct paths and multipaths between all necessary pairs of points taken from a vast set.

The best we can hope for -- and this it seems is realistic -- is to say with some chosen confidence, that if equipment is approximately where it is supposed to be, if the various environmental conditions are approximately those used as model inputs, if we have studied a sufficiently large sample of the equipments in question -- then a given communication link will exhibit a performance somewhere between points A and B on some scale.

It may have become apparent that the word validation is being used in the EMETF in two somewhat different contexts. One may be called validation for development. This consists of whatever tests or comparisons may be useful in checking out the development of a module, particularly the propagation, equipment and scoring modules, as implied in figure 3.

The IPM is designed to be a theoretical model rather than an empirical one. That is, it performs the calculations textbooks and research papers give in explanation of what happens. It is not supposed to store empirical data on what has been observed at various times and places, and regurgitate the solution to the stored problem which is most similar to the desired problem. While this theoretical approach is the cause of considerable grief during development, the advantages of a good theoretical model over an empirical one are evident.

However, as it happened, we could not wait the millenium without useful outputs from the facility. Our sponsors presently began to clamor for results, and results we had to produce, even though we knew that none of the modules performed to our standards. We were thus, in part, forced into interim empirical solutions. From this there arose another concept, validation for utilization, a statement about the goodness of our results. We shall henceforth be concerned here with this latter sort of validation.

A block diagram of the entire test problem, figure 4, will be a helpful introduction to the next section. This shows the test transmitter with its special test signal and the propagation path to the test receiver and scoring device described later, labelled VIAS. The figure also shows potential interfering transmitters (IG), each of which is supplied with normal modulation, i.e., voice, radioteletype, etc.

The two scales below the diagram are merely subjective guesses, and no precise quantitative interpretations should be given to them. For example, on the adequacy of representation scale, which applies to the IPM, we know that both the test signal and the test transmitter are represented more adequately than is the test receiver. This is because most of the various things which happen to a signal as it passes through the equipment occur in the receiver. We also note that the receiver in turn is more adequately represented than is propagation loss.

The bottom scale includes not only factors related to less than perfect representation in the model, but also includes variation in electrical characteristics among equipments, time-dependent variation over a propagation path, and the like.

However inaccurate these judgements may be, they did provide some guidance for separating the total problem into parts. One easy choice to make, and one which is also required by the operational scheme, consists of fragmenting the problem into the interfence versus the non-interference cases; that is, study the problem with and without the interfering transmitters activated. The balance of this paper will be concerned only with the non-interference case.

Another division point was taken at the input to the receiver. This was selected on a recent test for several reasons. One is, it appears from the diagram that the first portion of the chain, from test signal to receiver input, would be basically a measure of the ability of the model to predict propagation loss. Another reason is that better measurements can be made on the low level signals at the receiver than on the high level signals at the transmitter. A third is that in the workshop we studied in detail the receiver-VIAS subsystem, and here the input to the subsystem was of necessity the input to the receiver. Thus the non-interference case was divided into two segments.

From the field facility, data on the received signal can be obtained from a number of transmitters at one or more receiver sites, each transmitter-receiver combination defining a path. In the IPM, these paths may be simulated and the computer signal at the receiver obtained for each. Thus we generate a set of bivariate data as shown in figure 5.

These data were treated by the method of simple linear regression. Since the regression will be used to provide an interval estimate for the expected value of a hypothetical "field" signal given a model signal, the latter was used as the independent variate. The confidence band shown is that for the line as a whole. In other words, it is based on the tabular factor

$\sqrt{2F_{2,\,n-2}}$ rather than on t_{n-2}, which is valid for only one

prediction of the expected value for Y, given X. This confidence belt has roughly 25 percent greater width than the one of the same confidence level which is computed using the Student t. We used this method for showing how well, on the average, the IPM was predicting path loss.

The first specific question directed to the Panel arises here. Is there a method for providing an interval estimate for any number of individual predictions?

To lead into the second question, further details are helpful. The regression is based on the received signal measured in negative dbm, that is, in decibels below one milliwatt. This is a measure of the signal power induced across the input impedance of the receiver. The gain of the receiver antenna is thus included in the signal power measurement. In practice, however, it may be necessary to measure the field intensity, a voltage impinging upon the receiver antenna. Thus the dbm measure shown may contain a computed element. Most likely this would be computed once for each type of antenna-receiver combination, and would not include interantenna variability or variable ground plane effects. Thus the dependent variate may among other things contain a fixed, computed component rather than a measured, variable one. Clearly, we need in such cases to assess the effects on our predictions.

This and other examples not cited pertain to the general question of whether we do in fact satisfy the several assumptions inherent in the regression model, which now suggests the second question. If conventional regression, upon further examination, is not applicable, can the Panel suggest alternative approaches? Remember that for subject matter reasons we desire to obtain, at approximately this point in the chain of events, a measure of how well the IPM is performing its job.

I will close this section by pointing out that we fully realize the measure of our ability to predict path loss over one terrain type, in an area of sparse vegetation, is not necessarily indicative of how well the model will perform under other terrain-vegetation combinations. The Army and others are presently engaged in collecting propagation data in various areas of the world. At present, however, the Arizona data are all we have to work with. It is our hope that by the time suitable data from other areas are available, we will have established the techniques to use these.

Before proceeding to the ~~second portion of the~~ non-interference chain, it will be helpful to describe the scoring device. The Voice Interference Analysis Set (VIAS), is a commercial device designed to convert signal-to-noise type information from the terminal end of the receiver audio section into a measure which is monotonically related to intelligibility. The result is the Articulation Index (AI). The conversion is accomplished by subdividing the audio frequencies from 200 to 6100 cycles per second into 14 bands, each of which is supposed to contribute equally to speech intelligibility. In each band the signal-to-noise ratio is measured during 17 seconds of the 30-second test period. For signal-to-noise ratios of +18 db or higher, the signal-to-noise ratio is converted to unity; for ratios of -12 db or lower, the conversion results in zero. In between +18 and -12 db, the conversion is approximately a linear function of the signal-to-noise ratio. The final articulation index is simply the mean of the 14 increments. There are some additional manipulations involved, and a special test signal is required, but these details need not concern us here. This device is based upon studies by French and Steinberg, and by Beranek.

It should also be noted that if a voice communication system does not possess the full bandpass of 200 to 6100 cycles per second, the VIAS bounds the AI between zero and some value less than unity.

The second segment of the chain concerns the radio receiver and the scoring device. In the shop, a rather precise curve can be established which relates the signal power, developed across the receiver input impedance, to the AI output. This curve generally resembles that shown in figure 6. Three things should be noted. First, the figure shows hypothetical data and, if anything, the point scatter is excessive. An individual receiver produces data points which scarcely deviate from a smooth curve. Second, what information we have indicates that variation among receivers results primarily in a horizontal translation of the curve by no more than a very few db. This is apparently the result of variation in receiver sensitivity. Third, measurements which fall in the lowest fourth or fifth of the AI scale are difficult to make, and these exhibit a higher variability than those resulting from stronger signals.

It should also be stated that the effects of varying the modulation level at the transmitter are at present unknown in detail, but presumably are very important.

The mathematical nature of the AI/signal relationship is not known from theory, as far as we have been able to ascertain. An understanding of the manner in which the VIAS operates on the signal clearly explains the rounded corners. It also allows for a strictly linear portion in the descending leg of the curve, at least for some values of signal and noise. Finally, the bounds on the function are easily understood. Perhaps this is enough.

In practice, the probit transformation was applied to the AI axis and a reasonable linear trend was established. Although the potential applicability of the probit transformation is not immediately obvious, a study by our contractor indicated that it could be used. Confidence intervals were established by the methods appropriate to probit analyses, and then mapped through the inverse of the transformation to provide the confidence belt shown in figure 7.

The three curves to the right represent the fitted line and its confidence bands for a specific equipment type. The single line to the left shows only the fitted curve for another equipment type.

Note that the line on the left drops from maximum AI to zero over a spread of about 25 db, whereas the other curve takes a little over 50 db to drop to zero. When it is considered that the power output of the transmitters normally associated with these receiver types is in the vicinity

of +40 to +45 dbm, we see that a receiver over most of the possible range of signals is either performing at its best, or else is not extracting any intelligence whatever from the desired signal.

The use of the probit transformation was an expedient. It is clear that in some cases it is not appropriate, clear because the best fitting probit line obviously does not fit well. In particular, if there is a considerable segment of the curve which is linear in the descending region, the probit transformation is not suited.

The third question for the Panel is this: Please comment on the problem of providing both point and interval estimates for the functional relationship between input power and output AI.

An earlier topic, the scoring device, dealt with a conversion from an electrical measure, the signal-to-noise ration, to a psychoacoustical measure, aural intelligibility. The question naturally arises: Is the AI scale a suitable measure of intelligibility?

Previous work, notably by Kryter, had shown that AI was not linearly related to the articulation score (AS), where the latter is defined as the proportion of words recorded correctly by a listener. Kryter showed, further, that different AS/AI relationships were obtained depending upon the size of the word list. He and others have shown or suggested that such other factors as the type of noise, i.e., white noise, voice babble, or meaningful single voice interference, also affect the AS/AI transformation. We have recently verified that the electronic circuitry of the communication equipment also affects this relationship.

While there are some theoretical results which predict the functional relationship between AS and AI, our position is that, at present, the relationship must be established empirically. Naturally, we anticipate the day when the appropriate theory, to include parameter values, has been established and can be used to convert, in the IPM, from the last electrical-type measure to intelligibility.

The articulation score, as we use it, is defined as the mean proportion of correct responses given by five listeners to a transmission involving one of several 50-word phonetically balanced lists.

The experimental procedure requires that the word list be transmitted and recorded on magnetic tape. The transmission also includes the special test signal required for the AI measurement. For various reasons, we now imbed the test words in carrier phrases, and this procedure necessitates a transmission time of 16 minutes. Each tape is scored in a special listening facility by five operators. The AI signal is scored separately by the VIAS. Thus, one transmission produces one AS/AI datum point.

Figure 8 presents some recent AS/AI data acquired from different equipments. The actual points are shown only for the middle curve. The scatter of points shown is roughly typical of each curve.

These curves were supplied by our contractor. The center and lower curves are based on the Gompertz curve while the upper one is hyperbolic. The Gompertz curve,

$$Y = ab^{c^X}$$

with a taken as unity, has been given some theoretical justification by previous psychoacoustic studies. It was fitted in linear form by means of a log log transformation in which Y is 1/AS and X is AI. The transformation enabled simple linear regression techniques, including confidence bands, to be applied. The confidence bands were mapped through the inverse of the transformation to provide an approximate confidence belt for the line as a whole.

The next question to the Panel is doubtless now evident. Please comment on the problem of converting AI to AS.

In summary, we began with a complex total problem, restricted it to voice communication, and further restricted it to the non-interference case. The non-interference case has been broken into three segments, each treated as a regression. The dependent variate for one becomes the independent variate for the next, with the articulation score as the ultimate dependent variate. At present, approximate and conservative confidence limits can be placed on the expected value of the AS. We are aiming for the ability to place exact confidence limits on the individual AS predictions.

My final question to the Panel asks for discussion of this problem of ultimate interest.

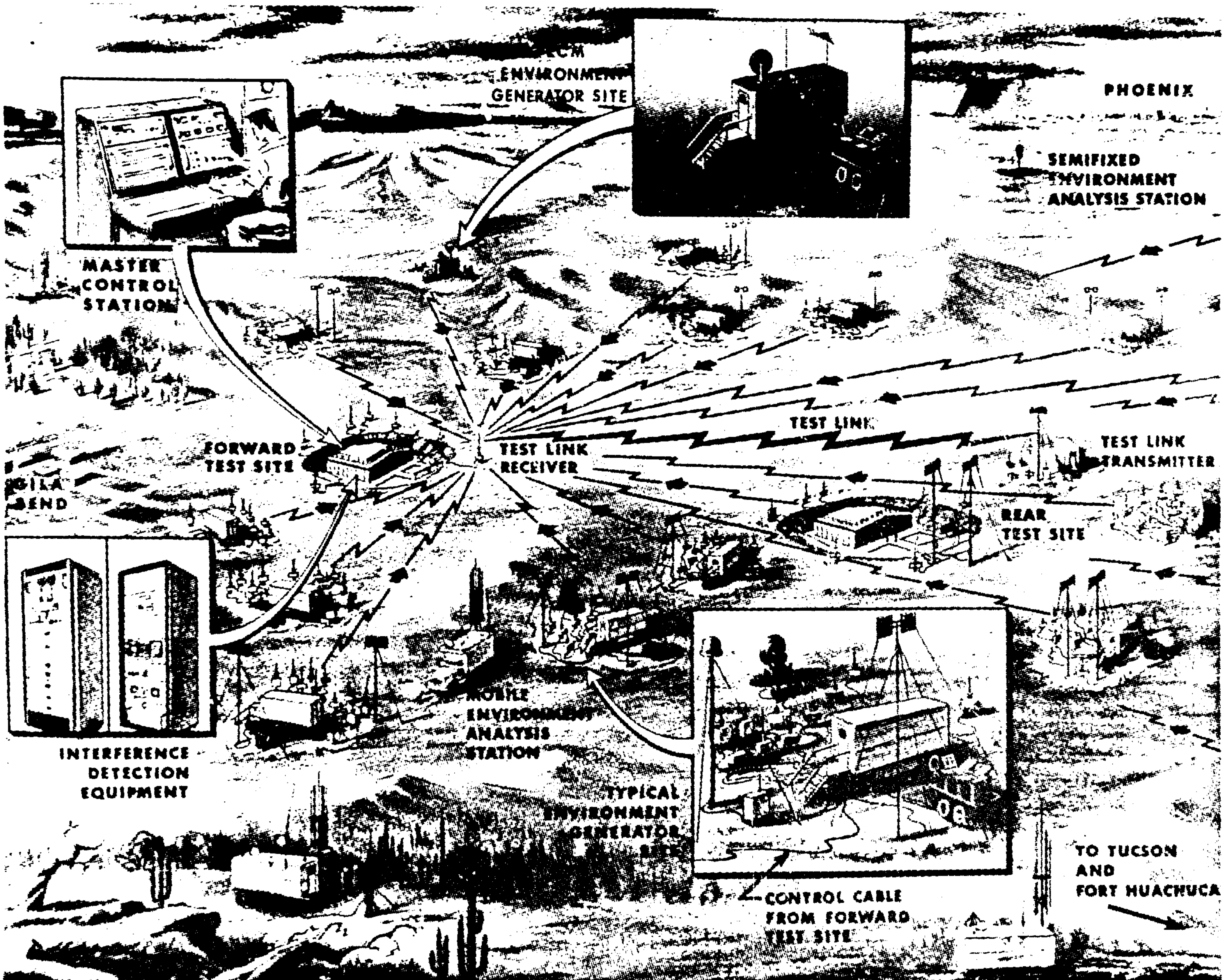


Figure 1. Artist's concept of the EMETF field facility.

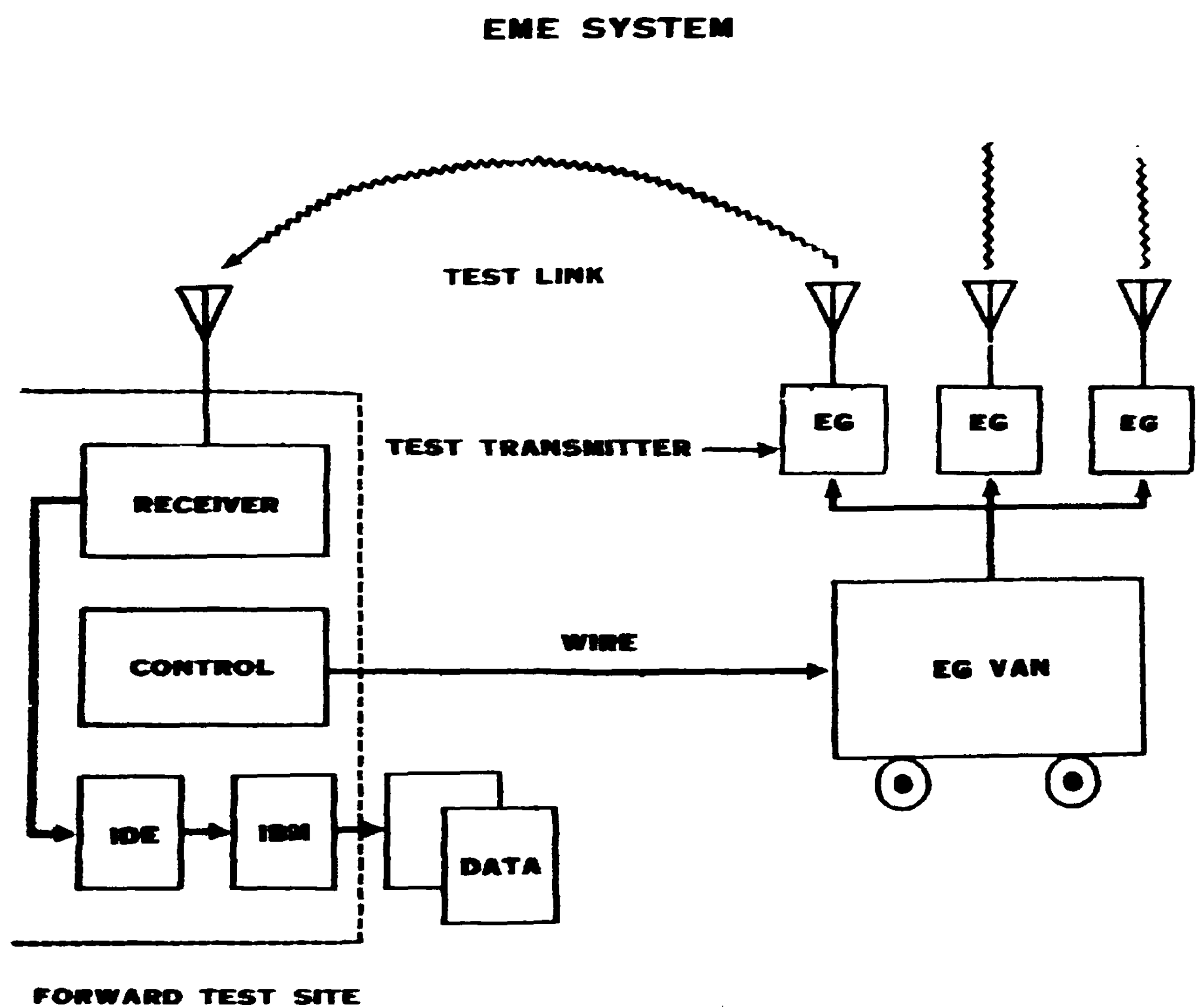


Figure 2. Test cycle of the EMETF

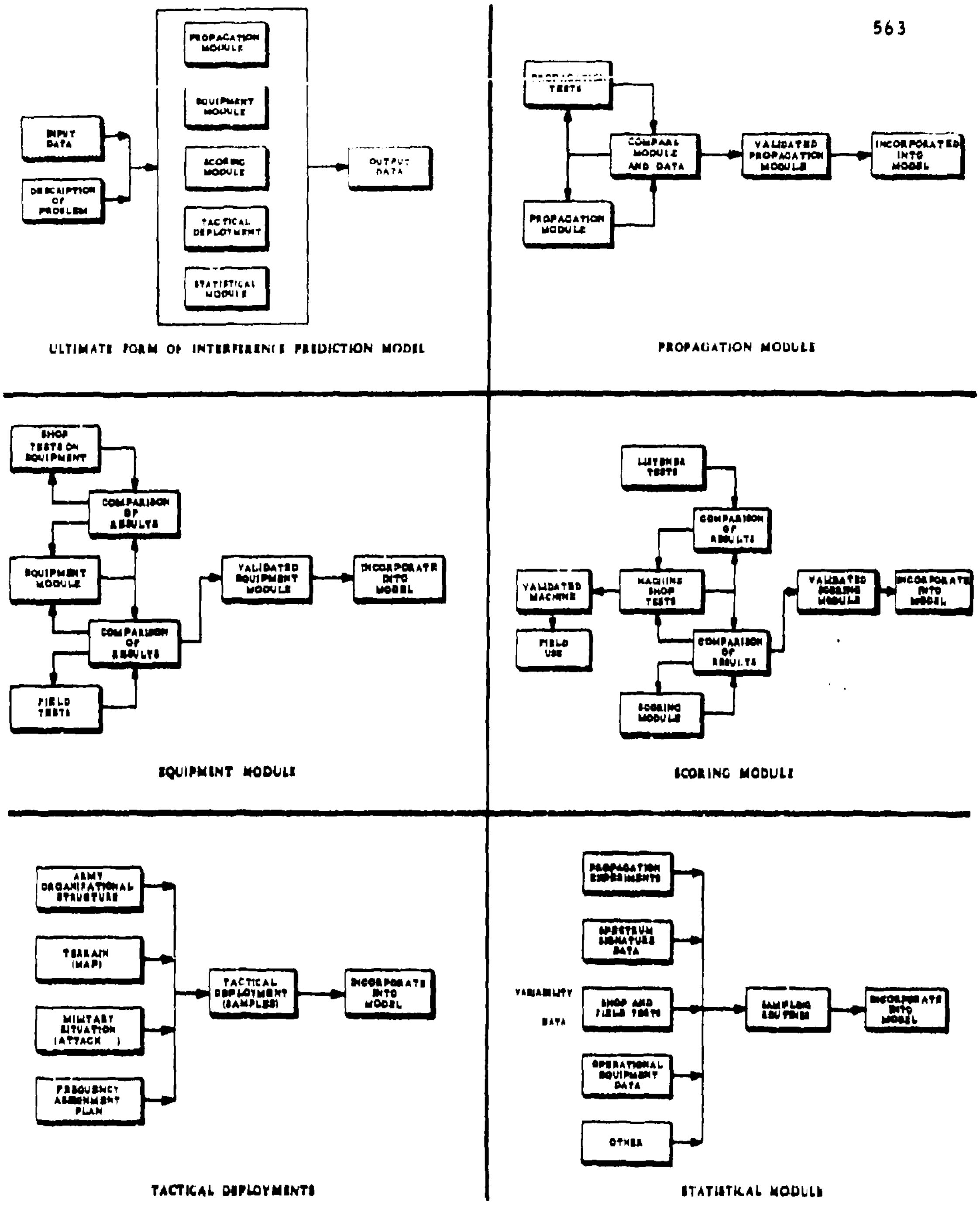


Figure 3. Block diagrams of the Interference Prediction Model and its Modules.

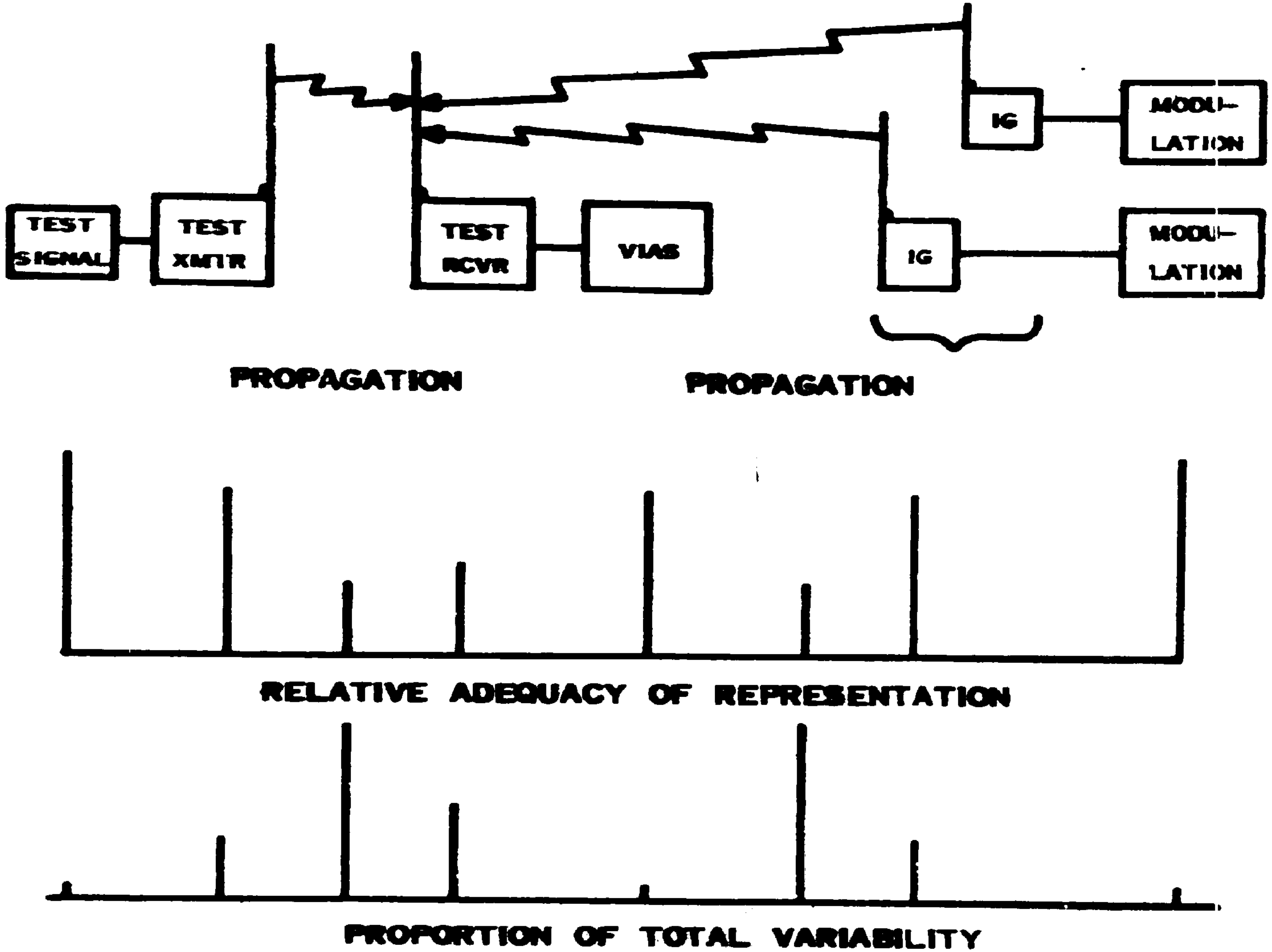


Figure 4. Block diagram of a single test.

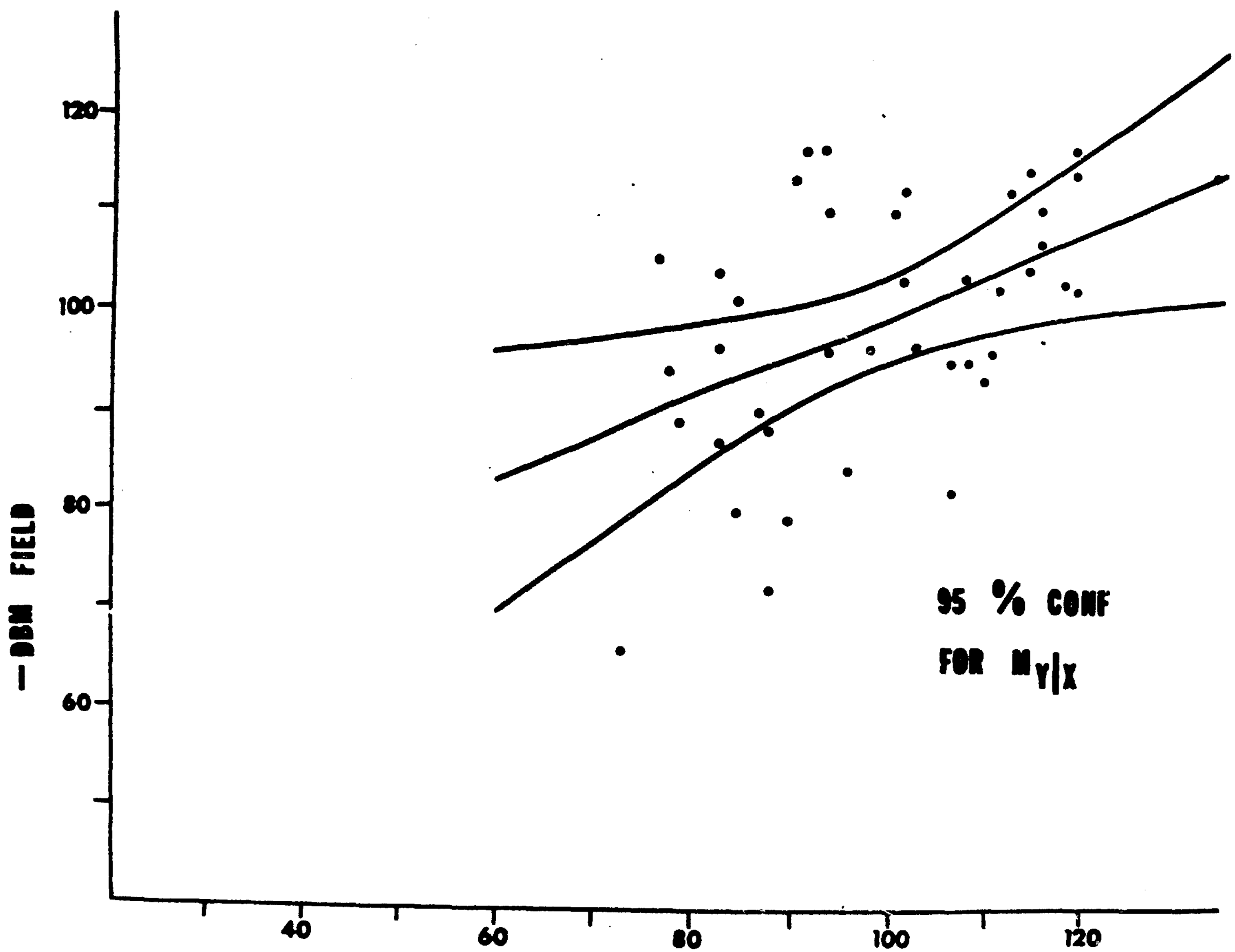


Figure 5. First section of the validation: Regression of received signal from field data against comparable model data.

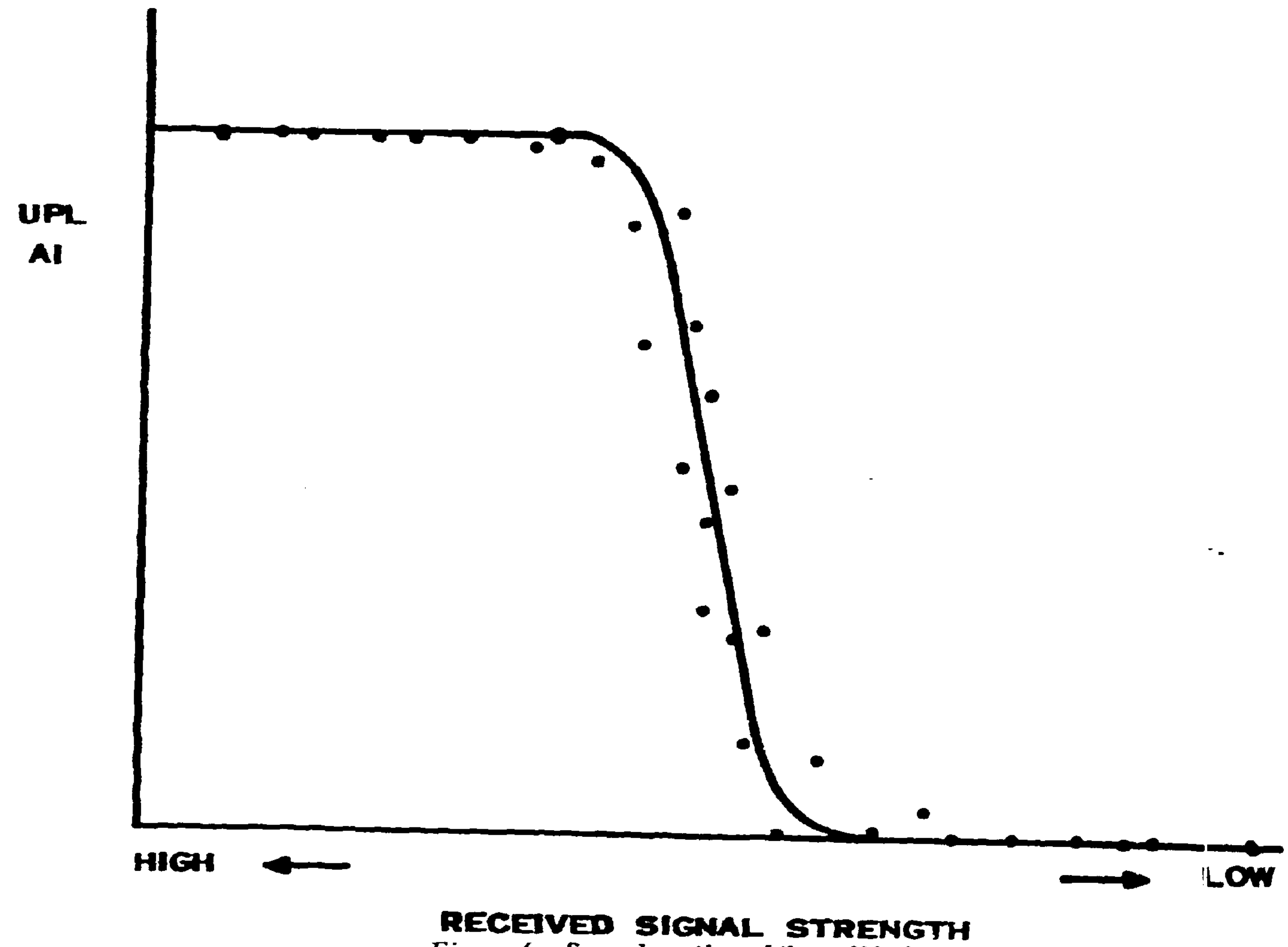


Figure 6. Second section of the validation.
Regression of AI against input signal. Hypothetical data.

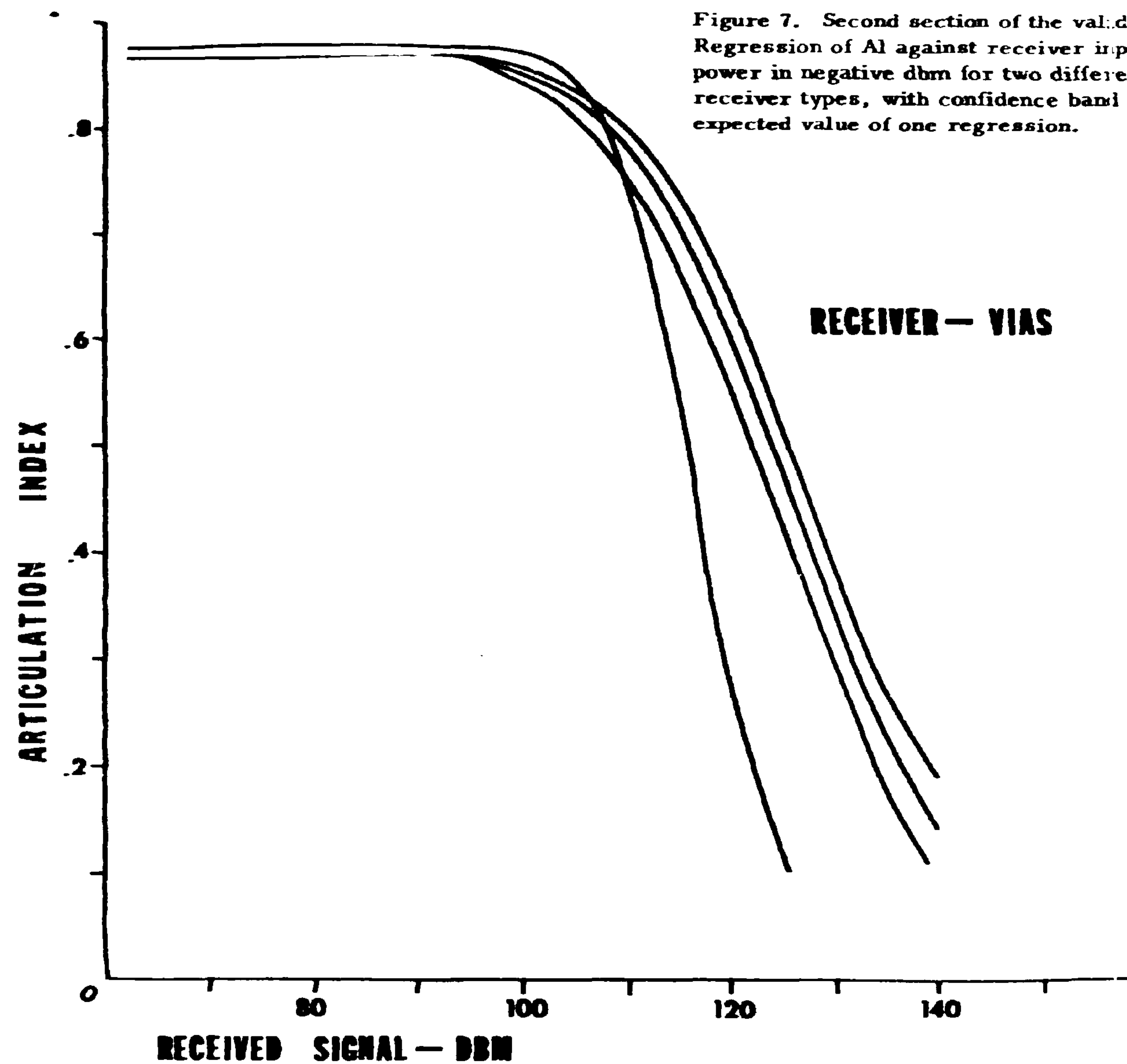


Figure 7. Second section of the validation. Regression of AI against receiver input power in negative dbm for two different receiver types, with confidence band for expected value of one regression.

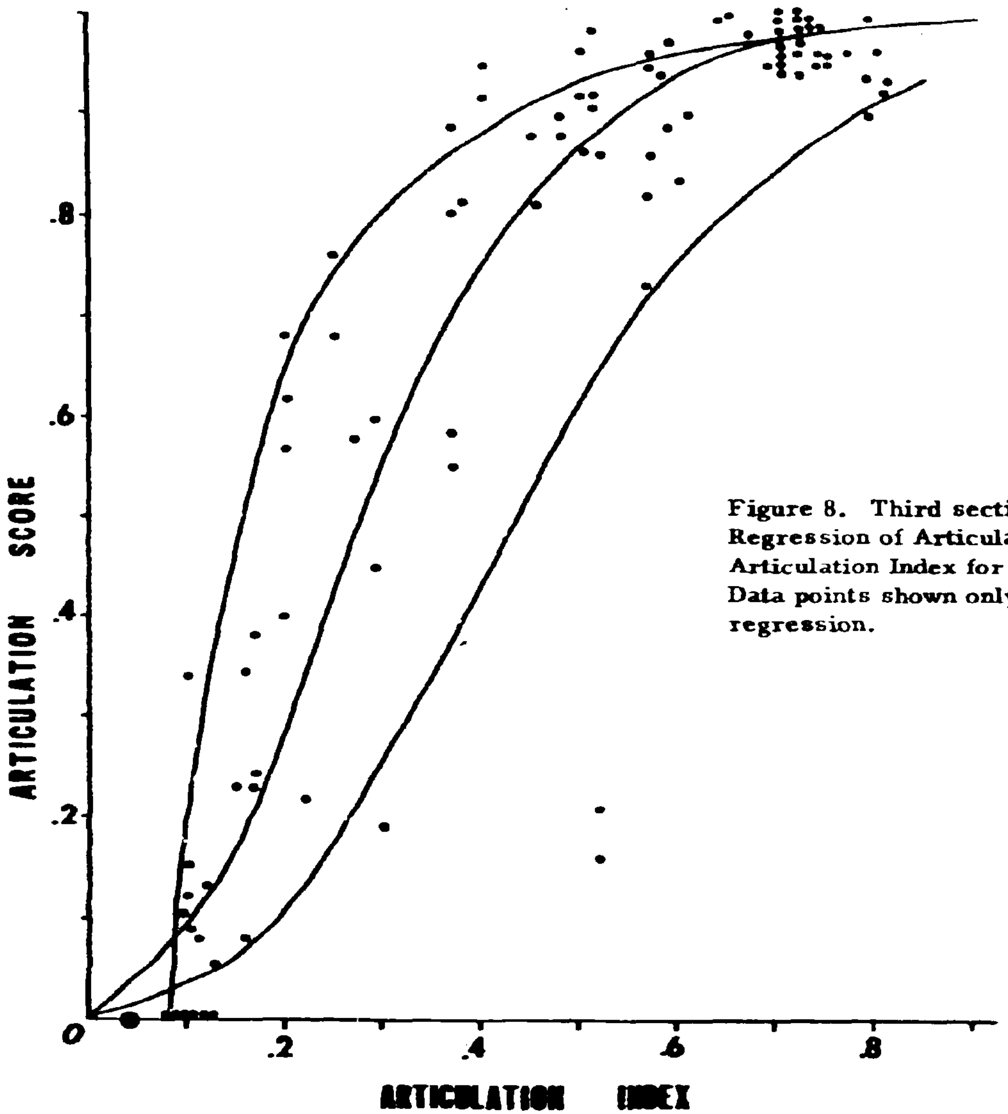


Figure 8. Third section of the validation. Regression of Articulation Score against Articulation Index for three receiver types. Data points shown only for the middle regression.

THE DESIGN OF COMPLEX SENSITIVITY EXPERIMENTS

D. Rothman and J. M. Zimmerman
Rocketdyne, A Division of North American Aviation, Inc.

1. INTRODUCTION. There is a growing tendency among the practitioners of the art of experimental design to allocate more of their efforts to the macroscopic aspects of test planning. This often results in greater benefit than that obtained from intensive improvement of isolated experimental segments. Very little work of this kind has been carried out for sensitivity experiments, however, despite the long history of statistical effort in this field, probably for two reasons. First, most of the major laboratories conducting sensitivity experiments have established over the years their own traditional set of test procedures which are relatively insensitive to variations in experimental objectives. Secondly, the majority of sensitivity experiments have been somewhat restricted in scope, being limited to such purposes as material screening or comparison of properties with those of a standard, and have not usually required extensive experimental planning and expenditures.

Recently it has become clear to many practitioners that there are several newer methods for the design and analysis of sensitivity experiments which deserve more substantial attention, partly because of their intrinsic merit and partly due to the increased complexity and cost of some current programs. It was in connection with one such program that the methods described in this paper were developed, although a substantial portion of the material had been previously formulated under a NASA, MSFC research contract, NAS 8-11061, monitored by Dr. John B. Gayle.

2. FORMULATION OF THE PROBLEM. Consider a sensitivity experiment in which there are n stimulus variables, $x_1, x_2, \ldots, x_n$, and for which the cost for each test is at least approximately known as a function of any combination of these variables. For simplicity, we assume that this cost is no different if the test response is positive (1) or null (0). Given α, suppose that the goal of the experiment is to estimate a specified portion of that $n-1$ dimensional surface S_α on which the probability of a positive response, $M(x_1, \ldots, x_n)$, equals α. Our analysis will be based on a loss function, L, which is made up conceptually of two terms: the cost of tolerating a specified variance in the

estimate of S_α, and the cost of testing. The overall problem is then to find that experimental design which minimizes $\overline{L}$, the value of L averaged over those portions of S_α which are of interest.

The treatment of the problem in this general form requires a carefully worked out technique for the design and analysis of multivariate sensitivity experiments which is readily amenable to the introduction of cost considerations. Although some algorithms for the design of multivariate sensitivity experiments have recently been developed (references 1 and 2), they are extremely complex and do not lend themselves to the implementation of loss minimization. Therefore, a simplification in the structure of the problem is required.

Towards this end, we replace the original multiple stimulus-variable problem by a hybrid regression-sensitivity problem in the following way. We select n-1 of the stimulus variables and consider them as independent variables in a regression model. The remaining variable (say the n^{th}) is considered as a stimulus with a possibly different response function at each combination of the n-1 regression variables. Effectively what we are doing here is replacing the n-variate response function $M(x_1, \ldots, x_n)$ by a univariate function $M(x_n; x_1, \ldots, x_{n-1})$ with parameters $x_1, \ldots, x_{n-1}$. Our program will be to estimate, at a set of specified values of these parameters, that value x_n^α of x_n for which $M(x_n; x_1, \ldots, x_{n-1}) = \alpha$; each point $(x_1, x_2, \ldots, x_{n-1}, x_n^\alpha)$ is in fact on S_α. Then we shall describe the effect of the parameters $x_1, \ldots, x_{n-1}$ by means of an ordinary regression of these variables on the estimates of x_n^α.

In a particular problem, the selection of the single stimulus variable from the original set is usually obvious, being dictated by the nature of the experimental apparatus, preparation of the test specimens, and long standing practice (e.g., in drop tests involving several environmental variables, such as temperature, orientation of the specimen, etc., the height of the impactor would invariably be the single stimulus chosen). In the present case, another important consideration which may affect the choice of the stimulus variable is the relative influence it has on the cost of testing. Our optimization procedure will be based only on the regression variables; that is, we determine the best combinations of

the variables $x_1, \ldots, x_{n-1}$ at which to test in order to minimize the loss function of the entire experiment. This macroscopic type of optimization does not itself dictate the local or microscopic design for the stimulus variable (x_n) at each of the regression parameter level combinations. Thus it is important in applying this method to select as the stimulus variable one which affects the cost of testing as little as possible.

It should be pointed out that this general approach to reducing the complexity of the problem is not new. For example, in 1961 Grant and Van Dolah described a procedure for handling multidimensional problems by the use of factorial designs combined with the simple up and down method (reference 3). In our work, however, the aspect of cost minimization has been added, and in addition a quantitative method for describing the efficiency of sensitivity experiments is developed. We treat these two topics in the following sections.

3. MACROSCOPIC COST OPTIMIZATION. The regression model relating the n-1 variables $x_1, \ldots, x_{n-1}$ with the estimates of x_n^a will be written in the form

$$x_n^a = P_o(\underline{x}) + P_1(\underline{x}) + \cdots + P_r(\underline{x}) + \epsilon, \tag{1}$$

where $\underline{x}$ is the vector $(x_1, \ldots, x_{n-1})$, $P_j(\underline{x})$ is a sum of terms of the j^{th} degree in the components of $\underline{x}$ with unknown coefficients, and ϵ is a normally distributed random variable with mean zero and (unknown) variance σ^2. Let N be the number of (not necessarily distinct) values of $\underline{x}$ at which test sequences on the stimulus variable x_n are to be run. The covariance matrix, Q, of the estimates of the coefficients in (1) can be written in the form

$$Q = \sigma^2 R(\underline{x})/N \tag{2}$$

where R is a matrix, independent of σ and N, whose elements involve averages of the components of $\underline{x}$ over the design. In treating particular

problems, one determines the optimum proportions of tests to be conducted at each of a certain fixed number of optimum treatment combinations, with N specifying the number by which these proportions are multiplied to obtain an actual design.

We have assumed that the average cost per test depends only on the vector $\underline{x}$ and not on x_n or N (it would depend on N if, for example, there were a setup cost). Let this cost be denoted by $C(\underline{x})$. For the moment we suppose that it is desired to obtain estimates of the function $x_n^{\alpha}(x_1, \ldots, x_{n-1})$ over an a priori specified region U with weighting function $W(\underline{u})$. Our loss function is a linear combination of the weighted average of the prediction variance over this region and the cost of testing. Thus the average loss is

$$\bar{L} = AN^{-1}\sigma^2 \int_U \underline{V}'(\underline{u})R(\underline{x})\underline{V}(\underline{u})W(\underline{u})d\underline{u} + BC(\underline{x}) \cdot N \tag{3}$$

where $\underline{u} = (u_1, \ldots, u_{n-1})$, A and B are appropriately chosen constants, and $\underline{V}$ is a column vector whose components are the linearly independent functions of the components of $\underline{x}$ contained in the quantities $P_j(\underline{x})$, $j = 0, 1, \ldots, r$ of equation (1). For example, in the very simple case when $\underline{x}$ is the scalar u and $r = 2$, we have

$$\underline{V} = \begin{pmatrix} 1 \\ u \\ u^2 \end{pmatrix} . \tag{4}$$

In this situation we have explicitly

$$Q = \sigma^2 \begin{pmatrix} N & \Sigma x & \Sigma x^2 \\ \Sigma x & \Sigma x^2 & \Sigma x^3 \\ \Sigma x^2 & \Sigma x^3 & \Sigma x^4 \end{pmatrix}^{-1} , \tag{5}$$

$$R = \begin{pmatrix} 1 & \overline{x} & \overline{x^2} \\ \overline{x} & \overline{x^2} & \overline{x^3} \\ \overline{x^2} & \overline{x^3} & \overline{x^4} \end{pmatrix}^{-1},$$

where $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$, etc. In the case when N (or the total cost) is not specified in advance we must find that value of N which minimizes (3). Since $R(\underline{x})$ is independent of N, on differentiating this expression one obtains

$$(6) \qquad N_{opt} = \sigma \sqrt{A \int_U \underline{V}'(u) R(\underline{x}) \underline{V}(\underline{u}) W(\underline{u}) d\underline{u} / BC(\underline{x})}$$

and the associated value of $\overline{L}$ is

$$(7) \qquad \overline{L} = 2\sigma \sqrt{AB \int_U \underline{V}'(\underline{u}) R(\underline{x}) \underline{V}(\underline{u}) W(\underline{u}) d\underline{u} \cdot C(\underline{x})} .$$

Thus, independently of the values of σ, A, and B, it is sufficient to minimize

$$(8) \qquad \overline{L}^2/4\sigma^2 AB = C(\underline{x}) \int_U \underline{V}'(\underline{u}) R(\underline{x}) \underline{V}(\underline{u}) W(\underline{u}) d\underline{u}$$

where the right hand member of (8) is proportional to the cost times the average prediction variance or "cost per unit of information". Note that this latter type of loss minimization may be accomplished independently of N and of the cost per unit variance ratio B/A. The value of $\sigma^2 A/B$ is explicitly required only if it is desired to determine N_{opt} from (6). If the total maximum expenditure of the test program is fixed in advance, as is often the case, then N is fixed and the values of σ, A, and B do not affect the minimization of the right member of (8).

When the region U over which the prediction variance is averaged is not specified a priori, the practical solution of the problem becomes more difficult. For example, it may be of interest in some problems to minimize the loss under the circumstances when an estimate of the value of $x_n^a(x_1, \ldots, x_{n-1})$ is to be made by extrapolation to a specified value of x_n^a, rather than to a given value of $\underline{x} = (x_1, \ldots, x_{n-1})$. In such cases, it is generally not possible to formulate the loss function explicitly in as simple a form as we have done since the coefficients in the model (1) are not known in advance. In this situation one may guess at the values of $\underline{x}$ at which the extrapolation is to be made and perform the optimization for a few such possibilities, or, alternatively, a formal Bayesian viewpoint can be taken, an a priori distribution of the extrapolation point made, and the optimization carried out formally in terms of this distribution. We will not pursue this more difficult version of the problem here, although it occurs not infrequently in practice.

When $r = 1$, and the form of the regression and cost models are simple, it is possible to carry out the minimization of (3) in closed form. However, the explicit optimum values of $\underline{x}$ are not always determined by this procedure. For example, we have shown (see reference 4) that when $r = 1$, there are p regression variables, and the cost function C is quadratic, then all that is specified by the minimization of (8) are the means and covariance matrix of the design variables. That is, the minimization of (8) provides

$$p\,\frac{(p+1)}{2} + p = \frac{(p+1)\,(p+2)}{2} - 1^*$$

constraints which the optimum design must satisfy. Now generally $k(p+1)-1$ constraints are required to define uniquely a design consisting of k distinct points. When $r = 1$, $p = 1$, for quadratic cost, we obtain $\frac{2 \cdot 3}{2} - 1 = 2$ constraints; this is one short of the $2(2)-1 = 3$ required for a unique two-point design. When $r = 1$, $p = 2$, we obtain 5 constraints; since three points are required to fit this model, thus requiring $3(3)-1 = 8$ constraints, we get a family of optimum three-point designs with three degrees of freedom.

*Alternatively we may say that, for quadratic cost, the number of distinct elements in the cross product matrix for the design, less one, gives the number of constraints obtained from minimization of (8).

When the cost function is made up of functions of components of $\underline{x}$ which do not already appear in the cross product matrix then one obtains an additional constraint from the minimization of (8). In fact, in the general case of any r and p we have made the following conjecture.

Conjecture: Let m be the number of distinct elements in the cross product matrix, P, corresponding to the polynomial model (1), of degree r. Suppose the cost function C contains functions of the components of the $p(=n-1)$ dimensional vector $\underline{x}$ which do not appear in P (we refer to this as condition I). Then minimization of the loss function (8) yields m constraints for the determination of the optimum design. If the cost only contains functions already appearing in P (condition II) then minimization of (8) provides $m-1$ constraints.

Since a design of k distinct "points" or treatment combinations requires $k(p+1)-1$ constraints for unique determination we have immediately the following:

Corollary: Minimization of (8) results in an optimum design consisting of

$$\max \left\{ q, \ \left] \frac{m+1}{p+1} \right[\right\} \quad \text{points when condition I prevails and}$$

$$\max \left\{ q, \ \left] \frac{m}{p+1} \right[\right\} \quad \text{points when condition II prevails,}$$

where $q = \Sigma_{s=0}^{s_0} \binom{r}{s} \binom{p}{s}$, $s_0 = \min\{r,p\}$; q = number of unknown parameters in the model, and $]\,y\,[$ denotes the smallest integer larger than or equal to y. The design is unique when the quantity in brackets is an integer. A formal proof of this conjecture may require solving the general minimization problem for (8), a very formidable task. Even the case $r = 1$ poses serious difficulties (see references 4 and 5). Apart from our verification of the conjecture in the linear case when condition II prevails (reference 5), we have recently solved a particular problem (using a computer search procedure) when U is a single point, $r = 2$, $p = 1$, and the cost function is exponential for all stimulus levels above a specified value. A unique three-point optimum design was found. Applying the conjecture and corollary (with condition I), we find in this case that the cross product matrix contains five distinct elements so that indeed five constraints are obtained determining a unique three-point optimum design.

In implementing this conjecture it is convenient to have the explicit relation between m, r and p. For example, for small values of the latter we have the following table.

p \ r	1	2	3
1	2	4	6
2	5	14	27
3	9	34	83

Values of m-1

$$m = \Sigma_{j=0}^{J} \binom{2r}{j} \binom{p}{j}, \quad J = \min(2r, p)$$

Thus, for example, if the regression model is cubic in three variables and condition II prevails, one would expect to find a unique 21-point optimum design. Note that in this case $q = 20$, so that the number of required points is greater than the number of unknown parameters.

Despite the formidable nature of an explicit closed form minimization of (8) in the general case, numerical minimization procedure may not require excessive effort. For example, the recently conducted study referred to above ($r = 2$, $p = 1$) only took a few minutes to run on an IBM 7094 computer.

4. BLOCKING OF THE TESTS AND THE GROWTH OF INFORMATION. Suppose we have obtained an optimum k-point design by the methods outlined above. The order in which these groups of tests are to be conducted is usually dictated by specific characteristics of the particular program. Generally the "least expensive" treatment combination or point (from the point of view of $C(\underline{x})$) will be explored first, then the next, and so on until the most expensive point is arrived at. We will not consider this question further here, but next turn our attention to the design of the individual group of sensitivity tests at each of the, say, q optimum treatment combinations of the regression variables.

Sensitivity experiments are most efficient when they are purely sequential, since in this situation one can reflect carefully on all previous

results before selecting the next test level for the stimulus variable. But if the experimenter is required for reasons of economy or manufacturing time limitations to order batches of test materials with specified (not necessarily equal) stimulus levels (as, for example, in solid propellant critical diameter studies), then it is necessary to consider the question of "block-sequential" sensitivity experiments and to evaluate the expected loss of information implicit in this mode of operation relative to the usual purely sequential test procedure.

To discuss block-sequential designs we will require a characterization of the amount of information available before the entire group of tests is conducted (from previous studies, etc.). This prior information will be expressed as that number of equivalent asymptotically optimal tests which would provide the same asymptotic information. Our approach will be based on the use of asymptotic expressions to characterize the growth of information in sensitivity experiments. Attention will be limited to the case in which the response function is a normal cdf, and to simplify the calculations we will assume that the sole aim of the tests is to estimate the median critical stress level (i.e., $\alpha = 50\%$). Our analysis will be carried out without actually specifying the test levels to be used in each of the blocks, although it is known (see reference 6) that for this type of experiment any test sequence converging to the median is asymptotically optimal in terms of efficiency in estimating the median. Evaluation of the validity of the asymptotic theory for small sample size is currently being studied by means of simulation.

Efficiency and Growth of Information. Suppose we have a cumulative normal response function with (unknown) parameters μ and σ. Let $\hat{\mu}$ denote the maximum-likelihood estimate of $\hat{\mu}$. Consider a design with T tests whose goal is the estimation of μ. An asymptotic expression for the variance of $\hat{\mu}$ (as $T \to \infty$) is given by (reference 7)

$$\sigma_{\hat{\mu}}^2 (T) \sim C_2 \sigma^2/(C_0 C_2 - C_1^2) \tag{9}$$

where y_i = the level of the stimulus variable on the i^{th} test,

$$
\begin{aligned}
t_i &= (y_i - \mu)/\sigma \ , \\
Z_i &= \frac{1}{\sqrt{2\pi}} \ e^{-t_i^2/2} \ , \\
p_i &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_i} e^{-u^2/2}\, du \ , \\
q_i &= 1 - p_i \ , \\
V_i &= Z_i^2 / p_i q_i \ , \\
C_j &= \sum_{i=1}^{T} V_i t_i^j \ .
\end{aligned} \tag{10}
$$

Since the goal of the experiment at each of the q optimum regression points is the estimation of μ, it is not unreasonable to restrict attention to designs which are asymptotically symmetric with respect to μ. Then

$$C_1 \sim 0 \ ,$$

$$\sigma^2_{\hat{\mu}}(T) \sim \sigma^2 / C_0 \ .$$

It has been shown (references 6 and 7) that $\sigma^2_{\hat{\mu}}(T)$ is asymptotically minimised when $t_i = 0$, $i = 1, \ldots, T$; this, minimum value is

$$\sigma^2_{\hat{\mu}}(T) \sim (\pi/2)\sigma^2/T \ .$$

The asymptotic information after T tests, I_T, may be expressed by the reciprocal of the variance of $\hat{\mu}$, or

(11) $$I_T \sim C_0/\sigma^2 = \sum_{i=1}^{T} Z_i^2/p_i q_i \sigma^2 \quad .$$

Thus the information contribution of a test at $t_i = t$ is given by

(12) $$I(t) \sim e^{-t^2}/2\pi pq\sigma^2 \quad .$$

Since this is maximized at $t = 0$, where we have

(13) $$I(0) \sim 2/\pi\sigma^2 \quad ,$$

the efficiency, defined as the relative information of an individual test at stimulus level t, may be written as

(14) $$E(t) = I(t)/I(0) \sim e^{-t^2}/4pq \quad .$$

The function $E(t)$ is tabulated below for selected values:

Table 1

$\|t\|$	0	.1	.2	.5	.75	1.0	1.5	2.0	3.0	4.0	5.0
$E(t)$	1	.9964	.9856	.9127	.8128	.6888	.4226	.2060	.0229	.00089	.000012

It may be noticed that the efficiency declines rapidly in the range $.75 < |t| < 2$. Tests for which $|t| > 3$ are very inefficient in the long run, although they may provide a large fractional increase in information early in the experiment.

In order to derive an expected value for $E(t)$, we express it in a more explicit form. It can be shown that the following expansions are convergent for all values of t:

$$p = 1/2 + t/\sqrt{2\pi} - t^3/6\sqrt{2\pi} + t^5/40\sqrt{2\pi}$$

(15) $$- t^7/336\sqrt{2\pi} + t^9/3456\sqrt{2\pi} - t^{11}/42240\sqrt{2\pi} + t^{13}/599040\sqrt{2\pi} - \ldots,$$

$$q = 1/2 - t/\sqrt{2\pi} + t^3/6\sqrt{2\pi} - t^5/40\sqrt{2\pi}$$

(16) $$+ t^7/336\sqrt{2\pi} - t^9/3456\sqrt{2\pi} + t^{11}/42240\sqrt{2\pi} - t^{13}/599040\sqrt{2\pi} + \ldots .$$

Therefore

$$pq = 1/4 - t^2/2\pi + t^4/6\pi - 7t^6/180\pi + t^8/140\pi$$

(17) $$- 83t^{10}/75600\pi + 73t^{12}/498960\pi - 523t^{14}/30270240\pi + \ldots,$$

and

$$E(t) \sim e^{-t^2}/(1-2t^2/\pi+2t^4/3\pi-7t^6/45\pi+t^8/35\pi$$

(18) $$- 83t^{10}/18900\pi+73t^{12}/124740\pi-523t^{14}/7567560\pi + \ldots) .$$

We have finally

$$e^{t^2}E(t) \sim 1 + 2t^2/\pi - 2t^4/3\pi + 4t^4/\pi^2 + 7t^6/45\pi - 8t^6/3\pi^2$$

$$+ 8t^6/\pi^3 - t^8/35\pi + 16t^8/15\pi^2 - 8t^8/\pi^3 + 16t^8/\pi^4$$

$$+ 82t^{10}/18900\pi - 304t^{10}/945\pi^2 + 68t^{10}/15\pi^3 - 64t^{10}/3\pi^4$$

$$+ 32t^{10}/\pi^5 - 73t^{12}/124740\pi + 1132t^{12}/14175\pi^2$$

$$- 356t^{12}/189\pi^3 + 704t^{12}/45\pi^4 - 160t^{12}/3\pi^5 + 64t^{12}/\pi^6$$

$$+ 532t^{14}/7567560\pi - 296t^{14}/17325\pi^2 + 599t^{14}/945\pi^3$$

(19) $$- 7808t^{14}/945\pi^4 + 48t^{14}/\pi^5 - 128t^{14}/\pi^6 + 128t^{14}/\pi^7 + \ldots .$$

In general, since μ and σ are unknown, we are uncertain as to just which value of t we are testing at; let this uncertainty be represented by a density $f(t)$ with mean M and variance U. Since we are trying to test at $x = \mu$ $(t = 0)$, and since $\hat{\mu}$ is unbiased and asymptotically normal, we have

$$M = 0 \tag{20}$$

$$U \equiv \sigma_t^2 = \sigma_{\hat{\mu}}^2(T)/\sigma^2 \quad . \tag{21}$$

Then the expected test efficiency is given by

$$\overline{E(t)} \equiv \int_{-\infty}^{\infty} E(t)f(t)dt$$

$$\sim \int_{-\infty}^{\infty} \frac{E(t)}{\sqrt{2\pi U}} e^{-t^2/2U} dt$$

$$= \frac{1}{\sqrt{2U+1}} \int_{-\infty}^{\infty} \frac{e^{t^2}E(t)}{\sqrt{2\pi U/(2U+1)}} e^{-t^2/2[U/(2U+1)]} dt$$

$$= \frac{1}{\sqrt{2U+1}} \int_{-\infty}^{\infty} \frac{e^{t^2}E(t)}{\sqrt{2\pi v}} e^{-t^2/2v} dt \quad , \tag{22}$$

using the substitution $v = U/(2U+1)$. Because of (19) this integral can be thought of as a sum (with coefficients given by (19)) of the even central moments, M_{2n}, of a normal distribution with variance v. We have

$$M_0 = 1 \; ,$$

$$M_{2n} = (2n-1)vM_{2n-2} \; , \tag{23}$$

from which it follows that $M_2 = v$, $M_4 = 3v^2$, $M_6 = 15v^3$, $M_8 = 105v^4$, $M_{10} = 945v^5$, $M_{12} = 10395v^6$, and $M_{14} = 135135v^7$. Then from (22) and (19) we have

$$\overline{E(t)} \sim \left[1 + \frac{2}{\pi}v + \left(\frac{12}{\pi^2} - \frac{2}{\pi}\right)v^2 + \left(\frac{7}{3\pi} - \frac{40}{\pi^2} + \frac{120}{\pi^3}\right)v^3 + \left(\frac{1680}{\pi^4} - \frac{840}{\pi^3} + \frac{112}{\pi^2} - \frac{3}{\pi}\right)v^4\right.$$

$$+ \left(\frac{83}{20\pi} - \frac{304}{\pi^2} + \frac{4284}{\pi^3} - \frac{20160}{\pi^4} + \frac{30240}{\pi^5}\right)v^5$$

$$+ \left(\frac{665280}{\pi^6} - \frac{554400}{\pi^5} + \frac{162624}{\pi^4} - \frac{19580}{\pi^3} + \frac{12452}{15\pi^2} - \frac{73}{12\pi}\right)v^6$$

$$+ \left(\frac{523}{56\pi} - \frac{11544}{5\pi^2} + \frac{85657}{\pi^3} - \frac{1116544}{\pi^4} + \frac{6486480}{\pi^5} - \frac{17297280}{\pi^6}\right.$$

$$\left.\left. + \frac{17297280}{\pi^7}\right)v^7 + \cdots\right] \Big/ \sqrt{2U+1}$$

$$\cong \left[1 + .636620v + .579234v^2 + .560060v^3 + .548604v^4\right.$$

$$\left. + .539890v^5 + .534223v^6 + .530582v^7\right] \Big/ \sqrt{2U+1} \quad . \tag{24}$$

This function is tabulated below for selected values:

Table II

U	0.0	0.2	0.5	1.0	1.5	2.0	3.0	4.0	6.0	8.0	10.0	15.0	20.0	40.0	100.0
$\overline{E(t)}$	1.000	.934	.853	.75	.68	.63	.55	.49	.42	.37	.34	.28	.25	.18	.11

For small values of U, the approximation given by the first two terms of the expansion,

$$\overline{E(t)} \approx 1 - .3634U \quad , \tag{25}$$

is sufficiently accurate. For large values of U we have $v \cong .5$ and numerical evaluation of equation (24) leads directly to

$$\overline{E(t)} \approx 1.132/\sqrt{U} \quad . \tag{26}$$

At this point we have only an asymptotic formula (24) for computing the expected efficiency of a new test, given $\sigma^2_{\hat{\mu}}$. But denoting by E_0 the prior information in terms of equivalent efficient tests, and by E_i the efficiency of the i^{th} test, we have from our definition of the information after T tests that

$$I_T \sim I_{T-1} + 2E_T/\pi\sigma^2 \quad . \tag{27}$$

Then one can show by an elementary induction that

$$I_T \sim \frac{2}{\pi\sigma^2} \sum_{i=0}^{T} E_i \quad , \tag{28}$$

or

$$\sigma^2_{\hat{\mu}}(T) \sim (\pi/2)\sigma^2 / \sum_{i=0}^{T} E_i \quad . \tag{29}$$

Equations (21), (24), and (29) can be used to asymptotically describe the growth of information in sensitivity experiments.

Of these equations, only (21) is exact. Equation (29) is asymptotically valid as $\Sigma_{i=0}^{T} E_i$ goes to infinity, which will happen if and only if E_0 and/or T become arbitrarily large. Equation (24) holds asymptotically on the $j+1^{st}$ test as $\Sigma_{i=0}^{j} E_i$ goes to infinity, in which case U goes to zero. But note that (24) and (29) do not give unreasonable results even for large values of U and for small values of E_0 and T. Thus we shall attempt to draw tentative conclusions even in the latter cases.

In the following example we see how slowly the individual test efficiencies increase in the course of a purely sequential sensitivity experiment. Note that we never have to specify the individual test levels in this line of reasoning.

Example. If $E_0 = .10$, then from (29), $\sigma_{\hat{\mu}}^2(0) \sim 15.71\sigma^2$. From (21), $U \sim 15.71$, and from (24), $\bar{\bar{E}}_1 \sim .275$. Continuing in this manner, we have $\sigma_{\hat{\mu}}^2(1) \sim 4.19\sigma^2$,

$$U \sim 4.19, \quad \bar{\bar{E}}_2 \sim .486,$$

$$U \sim 1.82, \quad \bar{\bar{E}}_3 \sim .647,$$

$$U \sim 1.042, \quad \bar{\bar{E}}_4 \sim .748,$$

$$U \sim .696, \quad \bar{\bar{E}}_5 \sim .811,$$

$$U \sim .512, \quad \bar{\bar{E}}_6 \sim .851,$$

$$U \sim .401, \quad \bar{\bar{E}}_7 \sim .878,$$

$$U \sim .328, \quad \bar{\bar{E}}_8 \sim .899,$$

$$U \sim .276, \quad \bar{\bar{E}}_9 \sim .911, \text{ etc.}$$

The above asymptotic theory has been tested by means of a computer program for simulating sensitivity experiments. The value of $\sigma_{\hat{\mu}}^2$ given by this theory is much more realistic than the value $(\pi/2)\sigma^2/T$, but still sometimes conservative by a factor of three.

Block-Sequential Designs. Now let us introduce the notion of a "block-sequential" design, in which each block of tests is planned after all previous test results have been analyzed. To "stage" such an experiment means to assign sample sizes to each of a given number of blocks, given the total sample size T. The "optimum" staging of a block-sequential sensitivity experiment is that staging which produces the greatest expected gain in information. Using the asymptotic methodology derived above, we

have computed a table of optimum staging for 2-block sensitivity experiments for total sample sizes up to 34 and for two different amounts of prior information.

Table III

Sample Size of First Block		1	2	3	4	5	6	7	8
Total Sample Sizes for which the Given First Block is Optimum	$E_0 = .02$	2-3	4-7	8-11	12-15	16-19	20-24	25-29	30-34
	$E_0 = 2.00$	2-3	4-6	7-10	11-14	15-19	20-25	26-31	32-34

For example, if 13 tests permitted were specified at the particular combination of regression variables under consideration, the optimum 2-block design would call for 4 tests in the first block and 9 tests in the second, over the given range of values of E_0.

It is fortunate that the above results are relatively independent of E_0, because this parameter is in practice very difficult to evaluate. For example, if our prior density on μ is uniform in $[A, B]$, then

$$\sigma^2_{\hat{\mu}}(0) = (B-A)^2/12 .$$

But to compute

$$E_0 = (\pi/2)\sigma^2/\sigma^2_{\hat{\mu}}(0) ,$$

we must know σ^2, and such information is almost always unavailable.

Results of the type given in Table III are not completely rigorous even for large values of E_0 and/or T, since we compute expected information in the second stage as a function of expected information in the first stage, rather than in terms of the distribution of this information. But the results are all plausible and of practical value precisely

because they are similar for E_0 = .02 and E_0 = 2.00. In addition, the optimum block sizes are obviously right for a total sample size of two, and the fraction in the first block decreases relatively smoothly as the total sample size increases.

It should be noted in passing that the above machinery permits us for the first time to characterise experiments in which the stress variable has an independent "setting" error, such as the projectile velocity in projectile penetration tests. Let this setting error be normally distributed with mean 0 and variance σ_s^2. Then the only change in the above formulae is in the expression for U, which is now

$$U = (\sigma_{\hat{\mu}}^2 + \sigma_s^2)/\sigma^2 .$$

Example: Let $\sigma_s^2 = 2\sigma^2$. Then the asymptotic efficiency of even a purely sequential design in this case is only 63%, since $U \sim 2$ (see Table II).

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of Dr. S. R. Webb in formulating the conjecture of Section 3. The computer simulation program was coded by P. M. Hait and B. J. Byars.

REFERENCES

[1] Madeline J. Alexander, "Multivariate Sensitivity Experiments with Non-Interacting Stimuli," Rocketdyne Research Report RR 64-15, July 1964.

[2] Madeline J. Alexander, "Models for the Analysis of Multivariate Experiments with Interacting Stimuli," Rocketdyne Research Memorandum RM 1121-351, 22 February 1965.

[3] R. L. Grant and R. W. Van Dolah, "Use of the Up-and-Down Method with Factorial Designs," Proceedings of the Seventh Conference on the Design of Experiments in Army Research, Development, and Testing, ARODR 62-2, pp. 39-65.

[4] "Statistical Design of Complex Experimental Programs, Final Report, I. Optimum Experimental Designs Obtained by Minimizing a Loss Function," Rocketdyne Report R-3392-1, March 1962.

[5] Madeline J. Alexander, N. R. Goodman, M. O. Locks, and S. R. Webb, "Some Results on Designs for Regression Experiments, Design of Experiments with Autocorrelated Errors Present, and Decision Theory Approach to Complex Experimentation," Aeronautical Research Laboratories Report ARL 63-107, June 1963.

[6] H. Chernoff, "Optimal Design of Experiments," Technical Report No. 82, Applied Mathematics and Statistics Labs, Stanford University, 31 October 1962, pp. 2-3, 8-15.

[7] D. Rothman, "The Inverse Response Problem for the Cumulative Normal Response Function, with Application to the Safety Problem," Rocketdyne Research Memorandum RM 1057-351, 28 July 1964.

FACTORS AFFECTING SENSITIVITY TESTING

James R. Kniss, and Warren S. Wenger
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

INTRODUCTION. Sensitivity testing is frequently utilized by the army in evaluating the sensitivity and consequently the reliability of percussion primers. Since such primers are used rather extensively in nuclear warheads, missiles and conventional munition systems, their functioning characteristics frequently have an important bearing on the reliability of these systems. However, knowledge of these characteristics of the primers and consequently of their reliability under varying temperature and impact conditions is often rather limited.

In a recent study involving the reliability of a nuclear weapon system, the effect of temperature and firing pin impact velocity on the reliability of initiation of a primer became of interest. This problem arose as a result of the procedures currently being used to test the system in which a particular type of primer is used. These procedures do not include a test of the primer itself, but the firing pin which initiates the primer is tested to determine whether the kinetic energy produced equals or exceeds a specified level of kinetic energy. The test results thus far obtained indicate that the level of kinetic energy specified is not compatible with the sensitivity of the primers. For, although a considerable number of firing pins have failed to produce the specified level of kinetic energy, in subsequent tests none of them has failed to fire a primer. Furthermore, the results of the primer testing that has been done indicate that the required kinetic energy is dependent upon the temperature of the primer. It has also been suggested that the sensitivity of the primers might not be a function of kinetic energy alone, but might also be a function of the impact velocity of the firing pin. If this relationship does exist then any primer test fixture should be designed to simulate the stroke velocity of the firing pin normally used to detonate that type of primer.

As a result of the questions that arose from this testing problem, a test was designed that would (1) measure the sensitivity of the primers under standard conditions, (2) determine the effect of strike velocity upon the kinetic energy required to function the primer, and (3) determine the effect of temperature upon the kinetic energy required to function the primer.

DISCUSSION OF TEST DESIGN. The test was to be conducted using the Bruceton Up-and-Down method of sensitivity testing. This method has been used for years, primarily in evaluating the quality or changes in quality of conventional primers. However, no record could be found of any test conducted where the strike velocity and temperature effects were investigated. Most past tests were conducted at ambient temperature and a single weight ball was dropped throughout. The height at which 50% of the primers would function was estimated along with the variability in this height. The results were used primarily to detect trends due to age and to detect lot-to-lot variability.

For this test the conventional procedures were modified in that four different weight balls and four different conditioning temperatures were used. A conventional type primer (the MK2A4) was used since the primer in question was not available and the MK2A4 is of a similar type. The primers were tested according to the following design:

		Ball Weight (oz.)			
		4	8	12	16
	70	X_{111}	X_{121}	X_{131}	X_{141}
		X_{112}	X_{122}	X_{132}	X_{142}
	25	X_{211}	.	.	.
		X_{212}	.	.	.
Temp. (^{o}F)	-20	X_{311}	.	.	.
		X_{312}	.	.	.
	-65	X_{411}	.	.	X_{441}
		X_{412}	.	.	X_{442}

Where the x's represent the drop height at which 50% of the kth sample of primers conditioned at the ith temperature and using the jth ball weight will function. However, it is obvious that the drop height will be affected by ball weight; and, of course, we are not interested in this obvious relationship.

In order to obtain the relationship in which we are interested, it will be necessary to convert these drop heights to kinetic energy using the following transformation

$$y_{ijk} = w_j x_{ijk}$$

where w_j is the ball weight in ounces and y_{ijk} is kinetic energy in inch-ounces. The kinetic energy obtained will be a function of the sensitivity of the primer and should be unaffected by the strike velocity or temperature if these factors are indeed insignificant.

It is, therefore, possible to hypothesize that if kinetic energy is the only factor affecting the sensitivity of the primer the analysis of the data should reveal no significant effects due to either temperature or ball weight. Should ball weight affect the required energy, it could be further hypothesized that this difference is due to the impact velocity of the firing pin.

At first glance the above design would suggest that a simple two-way classification of variables should be performed. However, it was suspected (and later confirmed) that the homogeniety of variances assumption which is necessary for this type of analysis might not be met.

This lack of homogeniety becomes intuitively obvious when it is considered that the change in kinetic energy per unit change in drop height is greater for the heavy ball. This would imply that the lighter balls yield better estimates of drop heights and consequently the variability associated with such estimates will be smaller for lighter balls.

If the data were analyzed as it is, erroneous results might be obtained as to the significance of the main effects as well as of any interaction that might exist between ball weight and temperature.

It was, therefore, planned to break the results down and first work with ball weight vs. kinetic energy at each level of temperature.

This, of course, does not solve the problem of the lack of homogeniety, and it is necessary to correct for this condition before progressing further. This may be accomplished by computing the within cell

variation for each ball weight and attempting to obtain the standard deviation as a function of required kinetic energy over columns, (ball weights) i.e.:

if we can obtain a relation $\sigma = f(\mu)$

then $\int \frac{1}{f(\mu)} \cdot d\mu$ is an appropriate

transformation that will transform the data so that the variability will be independent of the ball weight.

It is now possible to determine the relationship of ball weight to kinetic energy through the use of a simple least square analysis performed on the transformed data. However, it should be understood that the transformed data should be used only for purposes of significance testing and that the actual relationships should be represented by the un-transformed data.

It must be determined whether this relationship differs for each or any of the temperatures. If it does differ, is the difference only in intercept, only in slope, or in both intercept and slope? If it differs only in intercept the differences are constants over ball weight and all the data may be corrected back to ambient temperature so that a final relationship may be represented at ambient temperature (or at any other temperature within the range of the test that may be of interest). If the relationship differs in slope, an interaction between temperature and ball weight is indicated, and the relationship will necessarily be represented for each temperature or range of temperatures over which the slopes are homogeneous.

In any case a final representation of required kinetic energy (to function 50% of the primers) will be obtained and will be of the form $E = a + bw + cw^2 + \ldots$ (since a simple least square fit is being used). However, the relationship of kinetic energy (E) to strike velocity is of primary interest; and, therefore, the ball weight (w) in the above equation must be converted to strike velocity (v).

This may easily be done since a direct relationship exists between weight and velocity for any free falling body (neglecting air resistance, etc.). For example, assuming a linear relationship between required

energy and ball weight, the relationship between required energy and strike velocity may be obtained as follows:

Given the relation:

(1) $E = a + bw$

and the equations for free falling bodies:

(2) $S = (1/2)gt^2$

(3) $V = gt$

(4) $E = WS$

where E = Energy (in.-oz.)

a, b = constants

S = Drop height (in.)

g = acceleration of gravity (384 in./sec.^{2})

t = time (sec.)

V = Strick Velocity (in./sec.)

Solving equation (2) for t: $t = \sqrt{2S/g}$ substituting this value for t in (3):

$V = gt = g\sqrt{2S/g} = \sqrt{2gS}$

then substituting for S: $S = E/W$ gives

(5) $V = \sqrt{2gE/W}$

and solving (5) for W: $W = 2gE/V^2$

finally substituting in (1) and solving for E

$E = a + b(2gE/V^2)$

gives

$$(6)\quad E = \frac{a\,V^2}{V^2 - 2bg}\,,\quad \text{the desired relation .}$$

The final relationship of kinetic energy to temperature may be obtained in a similar manner. However, in this case there is no reason to believe that the variances will not be homogeneous.

DISCUSSION OF TEST RESULTS. In order to conduct this test, 32 samples of MK2A4 primers were fired, each sample being comprised of 40 primers. Using the Bruceton up and down method, the 32 estimates of 50% points were as follows for each combination of temperature and ball weight used:

50% Points (in inches)
Computed from Up and Down Tests*
of Primer, Percussion, Mk2A4

		Approximate Ball Weight (oz)			
		4	8	12	16
Temp. (°F)	70°	8.7500	5.3289	3.1375	2.4375
		9.4167	5.4250	3.7829	3.0329
	25°	9.7361	4.8289	3.5461	2.6118
		9.4342	5.2500	3.6591	3.1125
	-20°	10.0000	5.5395	3.5417	2.7875
		9.5000	5.6310	3.4934	3.1125
	-65°	9.8000	5.4868	3.9500	3.1000
		10.0109	5.5500	3.6500	3.4539

*A sample of 40 primers was used for each of the 32 tests.

Obviously there is a correlation between ball weight and the 50% points of drop height. But, of course, this is not very useful.

To get a meaningful basis for comparison, these heights were converted to the equivalent values of kinetic energy by multiplying each height by the corresponding ball weight (exact). Therefore, all further analyses were performed using the following values of kinetic energy:

Kinetic Energy (in. -oz.)
for above 50% Points

		Approximate Ball Weight (oz)			
		4	8	12	16
Temp. (°F)	70°	34.5562	42.4163	39.6570	39.7800
		37.1892	43.1813	47.8146	49.4969
	25°	38.4506	38.4365	44.8216	42.6246
		37.2583	41.7883	46.2498	50.7960
	-20°	39.4928	44.0926	44.7659	45.4920
		37.5182	44.8210	44.1555	50.7960
	-65°	38.7029	43.6732	49.9267	50.5920
		39.5358	44.1762	46.1348	56.3676

Since we suspected that the variances within ball weights might not be homogeneous, the individual cell variances were calculated and tested for homogeneity. This test not only confirmed our suspicions, but indicated a rahter acute case of non-homogeneity.

Further investigation showed that the relation between the standard deviation and ball weight could be satisfactorily represented by a function of the form $\sigma = a + bx$. And the required transformation, to correct for the observed non-homogeneity was found to be $y = 2.63 \ln(-13.8 + .38x)$ i.e. by substituting the E_{ijk} above for x in this equation we obtained y_{ijk} with homogeneous variances.

Having obtained these y's, we could then proceed to "determine" the relationship between ball weight and kinetic energy for each of the four test temperatures. A graphical representation of these relationships, together with the data from which they were derived, is given in Figure I.

The dots, of course, represent the data points and the lines the linear relationship derived from these points using least squares methods. Tests, using the transformed data (variances homogeneous), showed the slopes of these lines to be significantly different from zero, i.e., the 50% points of kinetic energy are a function of the ball weight used.

Visual comparison (see Figure II) indicated and a test confirmed that the slopes of these lines did not differ significantly, i.e., there was no reason to believe that the difference in the 50% point resulting from a given difference in ball weight varied with temperature. Or, to state it differently, the results of our analysis did not contradict the hypothesis that the difference in the 50% point resulting from a given difference in ball weight is independent of temperature.

If we accept this hypothesis, it follows that a better estimate of the effect of ball weight on the 50% point of kinetic energy should be obtained by "correcting" the data for temperature and then using the resulting (32) points to obtain a single relationship. This was done, and we obtained the equation K.E. = 33.51 + .81W as our best estimate of the relationship between kinetic energy and ball weight at ambient temperature (see Figure III).

We then obtained the desired relationship, between Strike Velocity and Kinetic Energy, by inserting the values of a and b from the above equation (a = 33.51, b = 0.81) into equation (6). This relationship was determined to be: $E = 33.51\ V^2/(V^2 - 620.28)$. Figure IV shows a graphical representation of this relationship and the points obtained for each of the 32 samples. The velocity and kinetic energy values used in plotting the points shown were "corrected" for temperature.

The data was also analyzed to determine the relationship between the 50% points of kinetic energy and temperature. As before, the cell variances were tested for homogeneity, this time they passed, i.e., no evidence of non-homogeneity was found.

We could, therefore, proceed to determine the relationship between temperature and kinetic energy.

Again using least squares methods, we obtained the relationships (see Figure V) between the 50% points of kinetic energy and temperature for each ball weight. (Comparison between points for 4 oz. and 16 oz. makes it apparent why the test in the first case indicated non-homogeneity of variances.)

Comparison of the slopes (Figure VI) confirmed that they did not differ significantly. Thus again a better estimate of the effect of

temperature on the 50% points of kinetic energy should be obtained by "correcting" the data for ball weight and using the resulting (32) points to obtain a single relationship. This was done (with data "corrected" to 16 oz.) and we obtained K.E. = 48.375 - .012 t as our best estimate of the relationship between kinetic energy and temperature (see Figure VII).

To summarize: We found that, for the Mk2A4 Primer, both temperature and strike velocity had a significant effect on the 50% point of kinetic energy, i.e., the kinetic energy required to fire 50% of the primers is a function of the temperature of the primers and the strike velocity of the firing pin as well as of the sensitivity of the primer.

While only the Mk2A4 primer was tested, we would expect similar results for other percussion primers. (If we had reason to believe otherwise we would not have used the Mk2A4 for this test).

Therefore, we feel, these results indicate the desirability of considering the effect of primer temperature and firing pin strike velocity on the kinetic energy required by other primers to assure reliable performance. Also, the desirability in testing primers of simulating the strike velocity of the firing pin normally used to detonate the primer is indicated.

Further, one might infer that investigation of the effect of strike velocity should be considered for sensitivity testing in general.

Kinetic Energy vs. Ball Weight
at Various Temperatures
for Primer, Percussion, Mk2A4

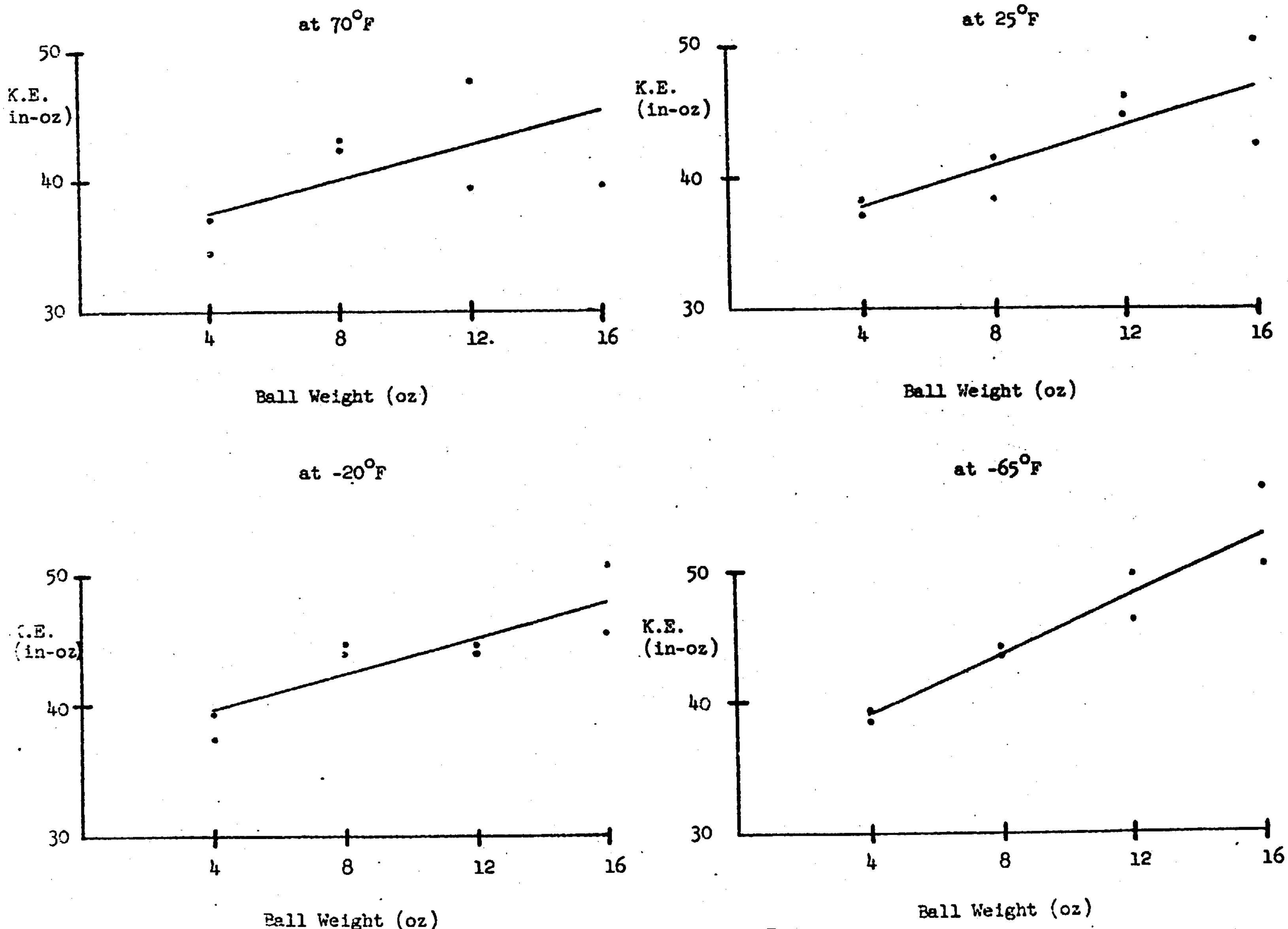


Figure I

Effect of Ball Weight on 50% Point
of Kinetic Energy - for Varicus Temperatures
(for Primer, Percussion, Mk2A4)

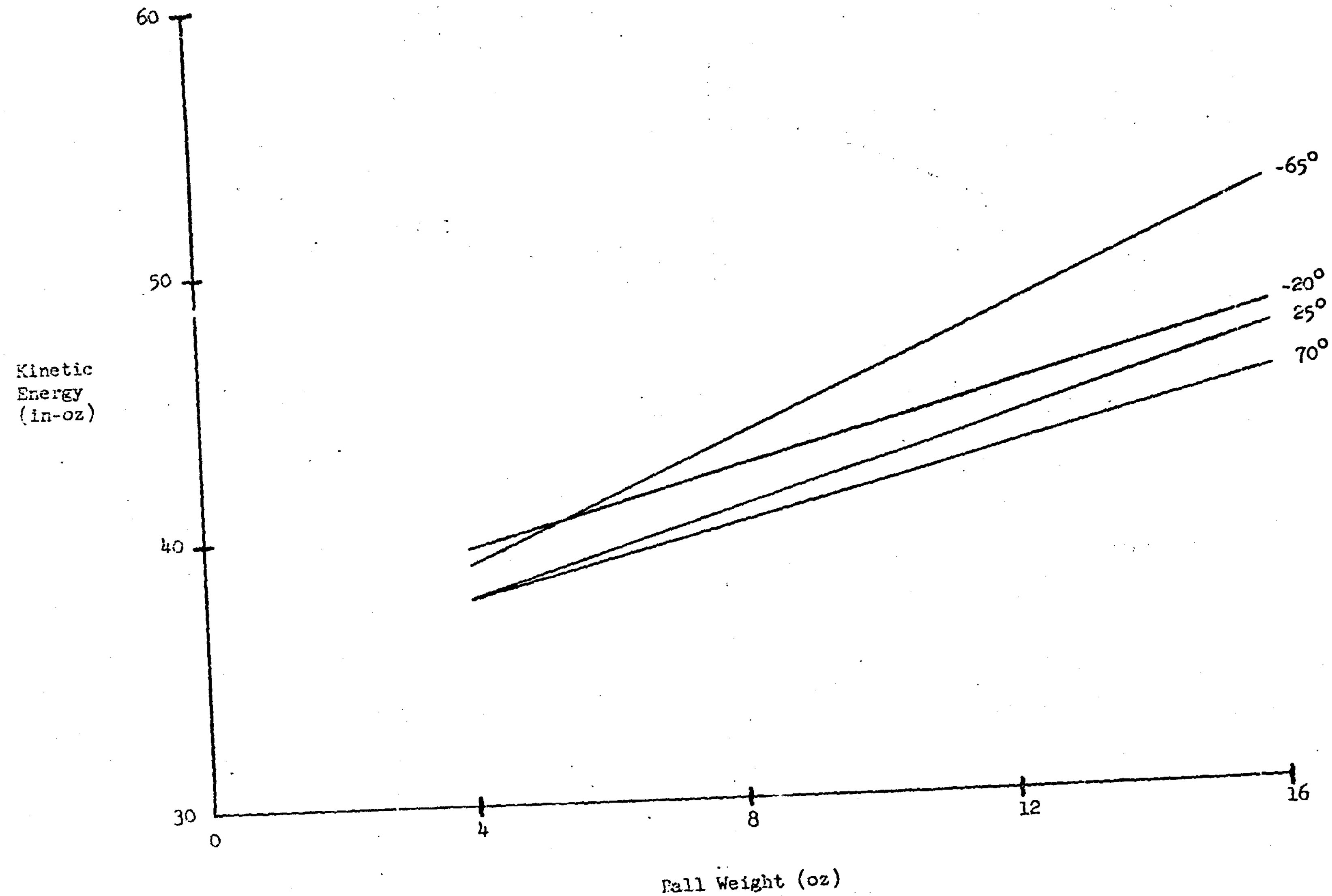


Figure II

Effect of Ball Weight on 50% Point
of Kinetic Energy
(for Primer, Percussion, Mk2A4)

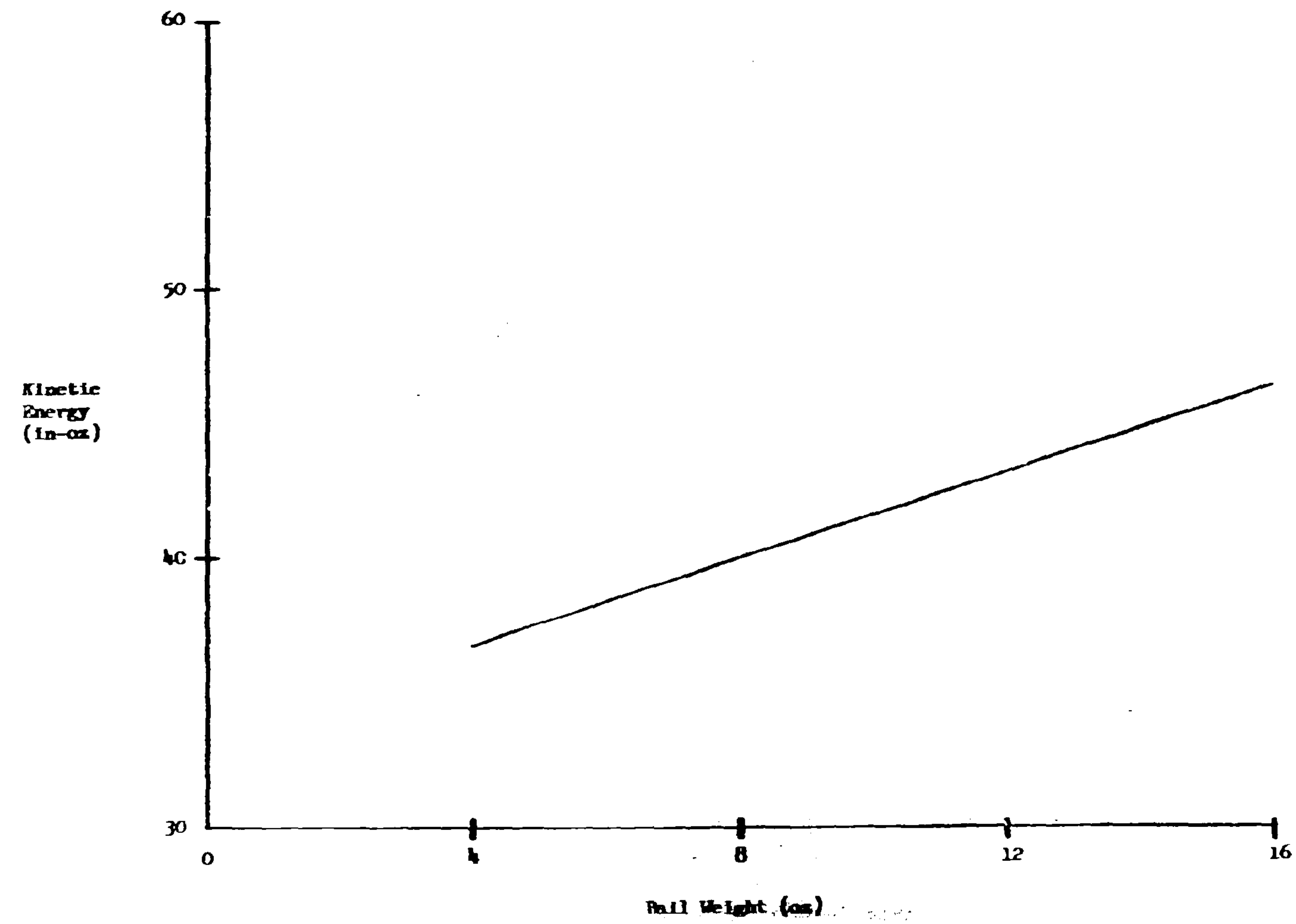


Figure III

Effect of Strike Velocity on 90% Point of Kinetic Energy
(for Primer, Percussion, Mk2A4)

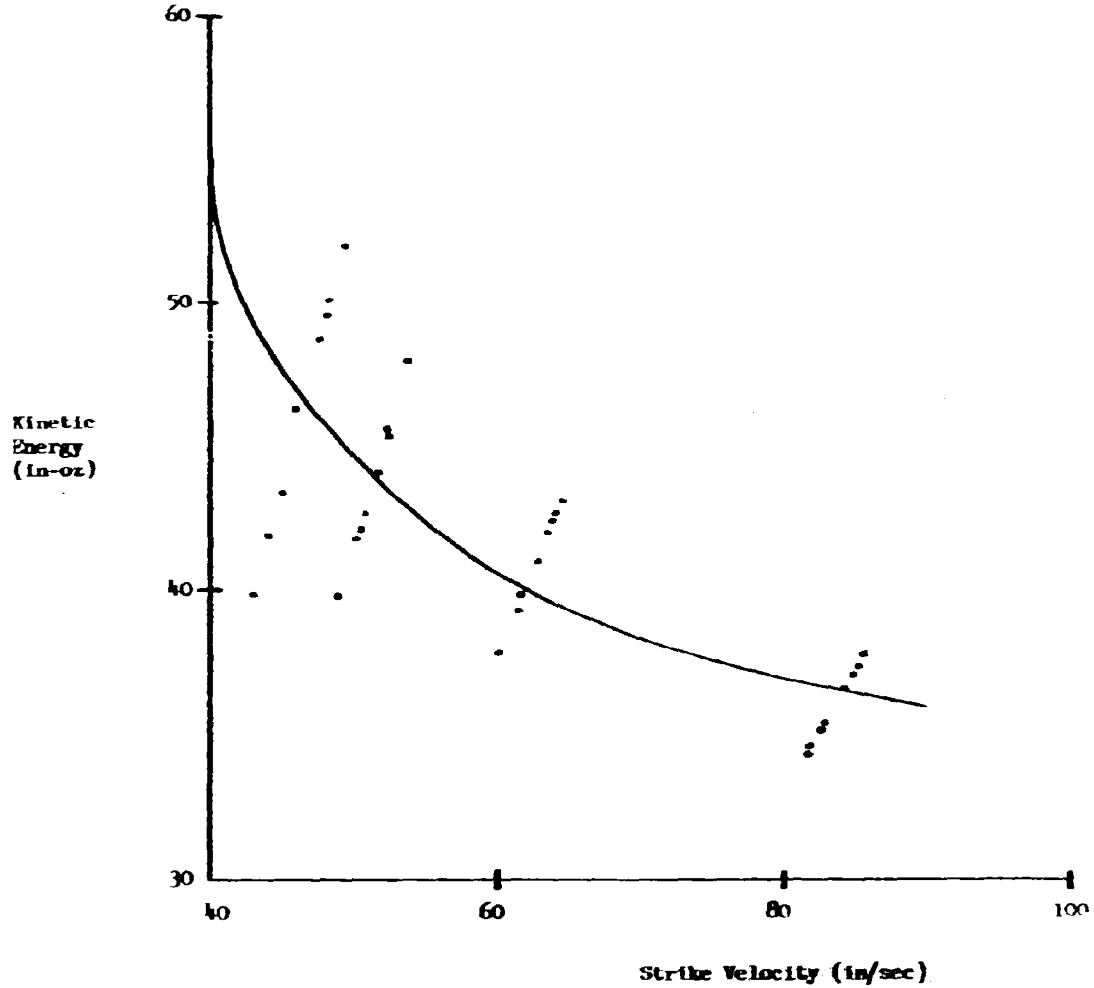


Figure IV

Kinetic Energy vs. Temperature
for Various Weight Balls
(for Primer, Percussion, Mk2A4)

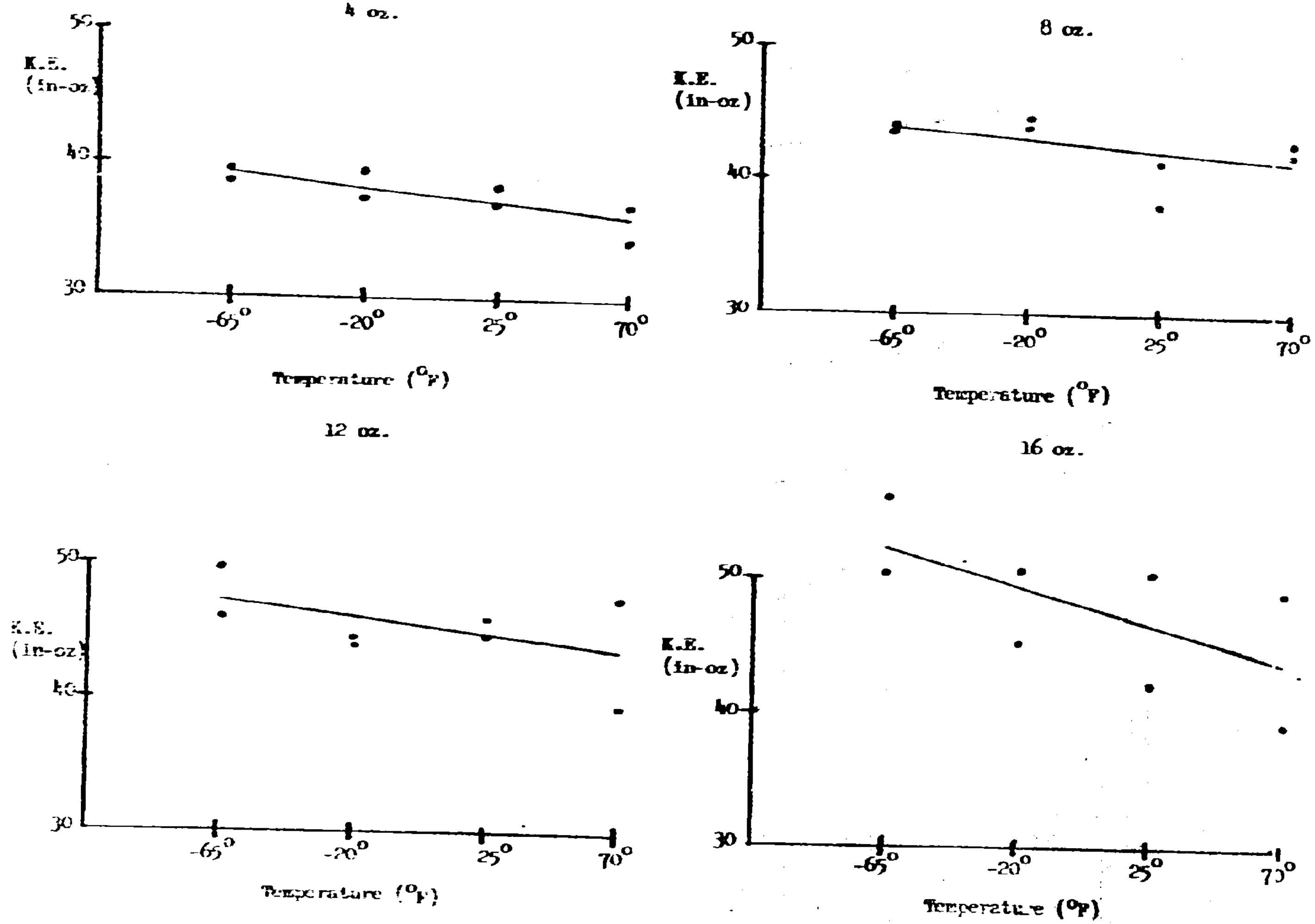
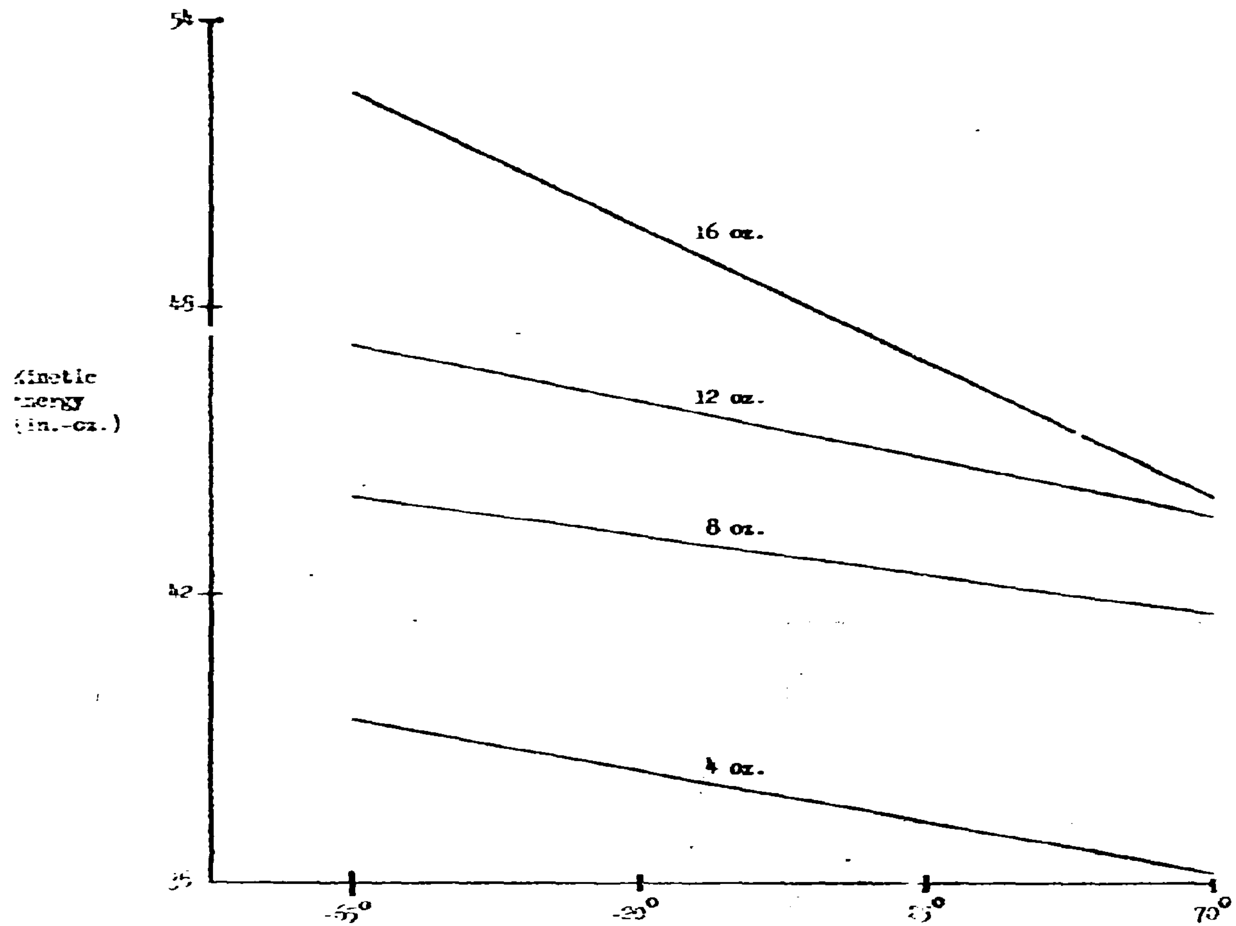


Figure V

Effect of Temperature on 50% Point of Kinetic Energy - for various Ball Weights (for Primer, Percussion, M29A4)



Temperature (°F)

Figure VI

Effect of Temperature on 50% Point
of Kinetic Energy
(for Primer, Percussion, M42A1)

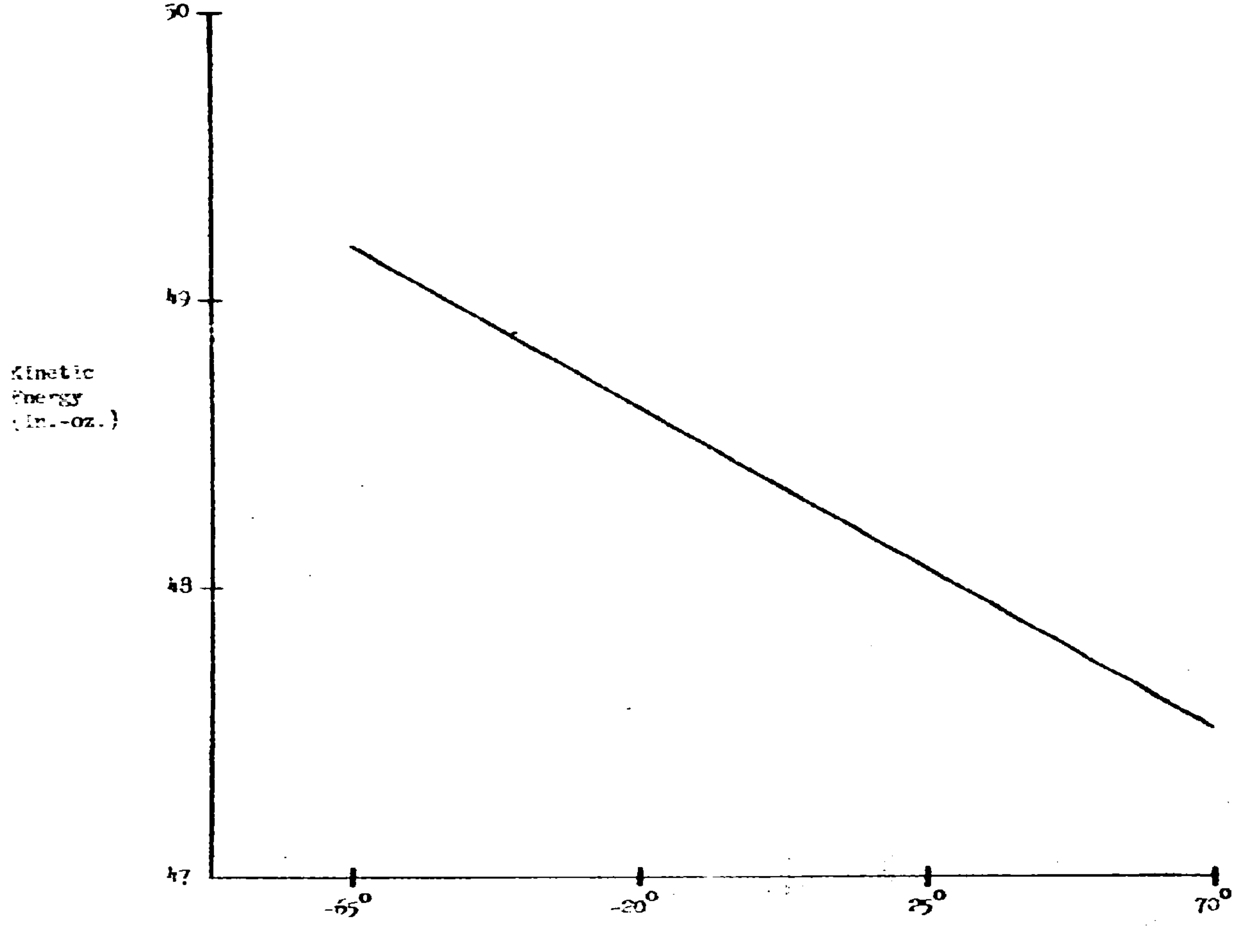


Figure VII

A COMPARISON OF RECONNAISSANCE TECHNIQUES FOR LIGHT OBSERVATION HELICOPTERS AND A GROUND SCOUT PLATOON

Harrison N. Hoppes, Barry M. Kibel, and Arthur R. Woods
Research Analysis Corporation, McLean, Virginia

INTRODUCTION. The Field Experiments Division of RAC is attempting to provide timely solutions to current Army problems involving tactics and doctrine. A major portion of the Division's field activities have dealt with helicopter operations. 1,2,3,4,5,6,7/ During July 1963 a research team from the Field Experiments Division conducted a two-sided, free-play field study with the 2nd Squadron, 4th Cav, 4th Armored Division to evaluate several techniques of helicopter reconnaissance. The results of that study were presented at the 9th Conference on the Design of Experiments in the paper "An Analysis of Helicopter Reconnaissance Techniques."

In November 1963, the Study's Project Advisory Group requested that a winter-phase investigation be carried out. The winter-phase venture measured the reconnaissance effectiveness of helicopters employing three reconnaissance tactics and compared the best of these tactics with the performance of a platoon of M114A1 Command and Reconnaissance Vehicles.

This paper describes the experimental design employed in the winter-phase investigation, summarizes the results obtained, and presents a brief statement of the study's conclusions and recommendations.

EXPERIMENTAL DESIGN. The experimental design is summarized in Table 1. As is indicated the three helicopter reconnaissance techniques studied were: (1) "high,"--flying at treetop level and maximum aircraft speed, (2) "low with pop up,"--nap-of-the-earth flight with emphasis placed on clearing an area before entering by popping up behind terrain masks, and (3) "low with dismount,"--nap-of-the-earth flight allowing the helicopter pilot to land and dismount an observer with binoculars. Single OH-13 helicopters, the vehicle currently used by the light-scout section of the air cavalary troop, 8/ were employed on all helicopter missions.

TABLE 1

Winter-Phase Experimental Runs

Ground Employment	Number of Runs For: Helicopter Tactic High	Low/Pop-Up	Low/Dismount	Ground Reconnaissance Platoon	Total Runs
Stationary	4	4	4	4	16
Moving	4	4	4	4	16
	8	8	8	8	32

The ground reconnaissance platoon generally consisted of five M114A1 scout vehicles. Usually the platoon leader divided the designated area or route into two sectors and coordinated the activity of the pairs of scouts operating in each sector. In performing their assigned mission, scout vehicle commanders frequently sent crew members forward on foot in much the same manner as helicopter pilots employed dismounted observers.

Like the companion study conducted during July 1963, the winter-phase investigation allowed scout elements complete freedom in determining paths of reconnaissance and time required to complete the assigned mission. Helicopter pilots were constrained only by the reconnaissance tactic they were instructed to employ; no restrictions whatsoever were placed on the ground reconnaissance platoon. Scenarios were designed to be tactically realistic and still permit experimental control.

Reconnaissance missions were conducted against static and fluid targets. Scout elements performed area reconnaissance missions against stationary target complexes and route reconnaissance missions against fluid targets. On each area reconnaissance mission scout elements reconnoitered against two target complexes, positioned to guard key terrain features and likely avenues of approach; each target complex consisted of one M113 APC and one or two M114A1's. On route reconnaissance missions target vehicles generally consisted of two M113's simulating the point of an armor column and three or four M114A1's providing route security. Target vehicles were mounted with gun cameras and event sequence recorders. Vehicle commanders were instructed to engage all reconnaissance elements acquired. Scout elements, on the other hand, were told to break contact whenever an enemy vehicle was acquired.

Throughout the paper the term "stationary runs" refers to those experimental runs involving stationary target complexes and "moving runs" to those involving fluid ground targets. Similarly, the term "target vehicle" is used to refer to the ground vehicles against which scout elements reconnoitered; it is never used to refer to reconnaissance vehicles taken under fire.

RESULTS. The winter-phase experimental design discussed above was successfully fulfilled between 20 January and 6 February 1964. A winter environment with snow cover, ground haze, and gray overcast was present on all days of field activity except February 5, 6.

Data, obtained from event sequence recorders and gun cameras mounted on target vehicles and from reconnaissance element sightings reported to a central control point, were analyzed using statistical techniques. Major emphasis was placed on comparing (1) the performance of helicopters vs ground scout teams, (2) the desirability of flying low/dismount vs high vs low/pop-up, and (3) the effects of reconnoitering against stationary vs fluid target complexes. The basic statistical technique used in making these comparisons was the analysis of variance; other common statistical techniques employed were t tests and chi-square tests.

Analyzing the results of two-sided, free-play experiments conducted in sector is often quite difficult. Frequently the outcomes of a given situation differ widely and the number of replications is small. At times experimental variables cannot be controlled as closely as is statistically desirable if troops and equipment are to be utilized when they are available. As a result, no attempt was made to analyze the experimental data in a rigorous manner. The statistical analyses did, however, provide an orderly framework for studying the large amount of data generated during the experiment.

Multiple measures of effectiveness were used in analyzing the experimental data. It was felt that no single measure could adequately consider all facets of the reconnaissance mission. Among the most important measures were those dealing with acquisitions, firings, and length of time required for mission completion. These included: (1) the percent of available targets acquired by reconnaissance elements, (2) the percent of ground targets acquiring at least one reconnaissance element, (3) the total number of times reconnaissance elements were detected, (4) the

number of times reconnaissance elements and ground targets saw each other first, (5) the average length of interacquisition advantages scored, (6) the percent of the time the reconnaissance element was heard before it was seen, (7) the average lay time against scout elements (8) the percent of reconnaissance elements acquired that were taken under fire by ground target vehicles, (9) the total number of individual weapon firings at reconnaissance elements (10) the number of simulated rounds fired at scout elements, and (11) the time required to complete reconnaissance missions. Each of these measures has its merits and its limitations. By considering a variety of measures the relative ability of reconnaissance elements to acquire targets, avoid destruction, and provide timely information can be estimated.

Summary data concerning these measures are shown in Tables 2-4. From these data it can be seen that:

1. Helicopters acquired about 60 percent of the available ground targets regardless of the reconnaissance technique employed. Based on the percent of ground targets acquiring a helicopter, the total number of times helicopters were detected, and the net number of acquisition advantages scored against helicopters, the low/dismount tactic was superior to the other two helicopter tactics examined.

2. Based on the number of firings and number of rounds simulated against helicopters, pilots employing the low/dismount tactic also outperformed those using the high and the low/pop-up tactics.

3. On the average it required 10 minutes to complete missions flying at treetop level and maximum OH-13 speed. Low/pop-up missions lasted twice this long and low/dismount missions $3\frac{1}{2}$ times as long.

4. The acquisition performance of a single OH-13 helicopter was quite similar to the performance of a platoon of five M114A1 scout vehicles. Both acquired about the same percent of available targets and both had 8 net acquisition advantages scored against them. Only according to one acquisition measure did low/dismount helicopters and the ground scout platoon differ widely; helicopters employing the dismount tactic were acquired audibly before they were seen 23 percent of the time compared with only 3 percent of the time for ground scouts.

5. On stationary runs the acquisition and firing measures listed in Table 3 indicated that the ground reconnaissance platoon outperformed the single OH-13 helicopter flying the dismount tactic. On moving runs, however, the low/dismount helicopter tactic was more effective than the ground scout vehicles.

6. In terms of all 11 performance measures summarized in Table 3, reconnaissance elements were more effective against the fluid target complex than against the stationary ground complexes studied. Many of the observed differences were quite large. For example, about twice as many acquisitions were made by stationary ground vehicles as by fluid vehicles, about 2½ times as many acquisition advantages were scored by stationary target vehicles as by moving targets, the mean interacquisition advantage against scout elements was twice as long for static units as for fluid, and over three times as many simulated rounds were fired by stationary vehicles as by moving vehicles.

CONCLUSIONS. Based on the summary data presented in Tables 2-4 and on the more detailed statistical analyses conduced for each effectiveness measure, it was concluded that in a winter environment against targets of the type studied:

1. The low/dismount tactic is more effective than the tactics of flying high or nap-of-the-earth with pop up.

2. The overall effectiveness of a platoon of M114A1 scout vehicles is similar to that of a single OH-13 helicopter employing the nap-of-the earth with dismount tactic. The ground scout platoon was more effective on the stationary runs and the helicopter dismount tactic on the moving runs.

3. The performance of both helicopters and ground scouts was significantly better against fluid vehicles with a movement mission than against stationary target complexes.

RECOMMENDATIONS. If it is decided to employ either a ground scout platoon or helicopters on winter-time reconnaissance missions in terrain similar to the type studied, it is recommended that:

1. Ground scouts be used against suspected stationary targets if time permits. If time does not permit, the helicopter tactic of low/dismount is suggested.

2. In reconnoitering against a fluid enemy, helicopters using the dismount tactic should be employed.

REFERENCES

1. RAC-TP-83, "A Preliminary Investigation of Helicopter-vs-Tank Operations (U)," February 1963. UNCLASSIFIED

2. RAC-TP-122, "Helicopter Operations with the Long Range Reconnaissance Patrol During Exercise BIG LIFT (U)," April 1964. SECRET

3. RAC-TP-124, "An Evaluation of the Laying and Tracking Capability of the M114A1 Scout Vehicle, Mainly Against Helicopters (U)." UNCLASSIFIED

4. RAC-TP-139, "A Preliminary Investigation of the Main Tank Gun vs Aerial Targets (U)." UNCLASSIFIED Technical Paper in preparation.

5. RAC-TP-136, "An Investigation of Gunner Tracking Ability Against OH-13 Helicopters employing Evasive Maneuvers (U)." UNCLASSIFIED Technical Paper in preparation.

6. RAC-TP- , "An Evaluation of Helicopter Pop-Up Tactics (U)." UNCLASSIFIED Technical Paper in preparation.

7. RAC-T-433, "Reconnaissance Techniques for Light Observation Helicopters in a Summer Environment -- A Two-Sided Field Play (U)." FOR OFFICIAL USE ONLY

8. Headquarters, Department of Army, FM 17-36, "Armored Cavalry Platoon and Troop, Air Cavalry Troop and Armored Cavalry Squadron (U)," December 1961.

TABLE 2

SUMMARY OF PERFORMANCE DATA FOR THREE HELICOPTER TACTICS, GROUND RECONNAISSANCE PLATOON

Effectiveness Measure	Helicopter Scout Tactic			Ground Scout Platoon
	Low/ Dismount	High	Low/ Pop-Up	
1. Percent of Ground Targets Acquired	59%	61%	55%	65%
2. Percent of Ground Targets Acquiring Reconnaissance Element	59%	81%	81%	51%
3. Number of Times Reconnaissance Element Acquired	31	57	51	51
4. Total Number of Acquisition Advantages Scored By Ground Targets	23	40	39	29
Total Number of Acquisition Advantages Scored By Reconnaissance Elements	15	9	13	21
Net Number of Acquisition Advantages Scored By Ground Targets	8	31	26	8
5. Length of Interacquisition Advantage Against Scout Element (Mean) in seconds	26	24	16	17
Length of Interacquisition Advantage Against Scout Element (Median) in seconds	10	21	10	20
6. Percent of Time Reconnaissance Element Was Acquired Audibly Before Visually	23%	21%	27%	3%
7. Lay Time Against Reconnaissance Element (Mean) in seconds	12	12	12	12
Lay Time Against Reconnaissance Element (Median) in seconds	10	8	11	8
8. Percent of Reconnaissance Elements Detected That Were Taken Under Fire	61%	58%	47%	72%
9. Number of Times Reconnaissance Elements Were Taken Under Fire By Ground Weapons	23	43	32	48
10. Number of Simulated Rounds Fired At Reconnaissance Elements	3424	5016	4880	2556
11. Time Required to Complete Mission (Mean) in minutes	35	10	20	93

TABLE 3

SUMMARY OF PERFORMANCE DATA FOR HELICOPTERS EMPLOYING LOW/DISMOUNT TACTIC, GROUND RECONNAISSANCE PLATOON

	Effectiveness Measure	Stationary Runs		Moving Runs	
		Helicopter Low/Dismount	Ground Scout Platoon	Helicopter Low/Dismount	Ground Scout Platoon
1.	Percent of Ground Targets Acquired	48%	62%	72%	68%
2.	Percent of Ground Targets Acquiring Reconnaissance Element	62%	62%	56%	41%
3.	Number of Times Reconnaissance Element Acquired	20	31	11	20
4.	Total Number of Acquisition Advantages Scored By Ground Targets	18	19	5	10
	Total Number of Acquisition Advantages Scored By Reconnaissance Elements	4	9	11	12
	Net Number of Acquisition Advantages Scored by Ground Targets	14	10	-6	-2
5.	Length of Interacquisition Advantage Against Scout Elements (Mean) in seconds	32	23	7	13
	Length of Interacquisition Advantage Against Scout Elements(Median) in seconds	12	23	5	10
6.	Percent of Time Reconnaissance Element Was Acquired Audibly Before Visually	43%	3%	10%	0%
7.	Lay Time Against Reconnaissance Element (Mean) in seconds	10	11	19	14
	Lay Time Against Reconnaissance Element (Median) in seconds	6	8	14	12
8.	Percent of Reconnaissance Elements Detected That Were Taken Under Fire	75%	68%	36%	80%
9.	Number of Times Reconnaissance Elements Were Taken Under Fire	18	28	5	20
10.	Number of Simulated Rounds Fired At Reconnaissance Elements	2984	1296	440	1360
11.	Time Required to Complete Mission (Mean) in minutes	49	117	21	70

TABLE 4

SUMMARY OF PERFORMANCE DATA FOR STATIONARY, MOVING RUNS

	Effectiveness Measure	Stationary Runs	Moving Runs
1.	Percent of Ground Targets Acquired	54%	68%
2.	Percent of Ground Targets Acquiring Reconnaissance Element	77%	57%
3.	Number of Times Reconnaissance Element Acquired	125	65
4.	Total Number of Acquisition Advantages Scored By Ground Targets	93	38
	Total Number of Acquisition Advantages Scored By Reconnaissance Elements	21	37
	Net Number of Acquisition Advantages Scored by Ground Targets	72	1
5.	Length of Interacquisition Advantage Against Reconnaissance Elements (Mean) in seconds	27	13
	Length of Interacquisition Advantage Against Reconnaissance Elements (Median) in seconds	13	10
6.	Percent of Time Reconnaissance Element Was Acquired Audibly Before Visually	26%	2%
7.	Lay Time Against Reconnaissance Element (Mean) in seconds	12	14
	Lay Time Against Reconnaissance Element (Median) in seconds	8	12
8.	Percent of Reconnaissance Elements Detected That Were Taken Under Fire	66%	48%
9.	Number of Times Reconnaissance Elements Were Taken Under Fire	106	40
10.	Number of Simulated Rounds Fired At Reconnaissance Elements	12,240	3736
11.	Time Required to Complete Mission (Mean) in minutes	52	26

A STUDY OF PROBABILITY ASPECTS OF A SIMULTANEOUS SHOCK WAVE PROBLEM

A method of solving probabilistic problems without a computer.

Edward C. Hecht
Picatinny Arsenal, Dover, New Jersey

I am going to present a procedure for the rapid solution by desk calculator of an involved probabilistic problem. Figure 1 is a sample computation, showing all the paperwork necessary for one solution. The first two and the last of even these few columns are identical for every computation of this sort.

Unfortunately, although characteristically, the evolution of the simple tool requires a long explanation. I have made up a problem as a vehicle for the explanation, and I hope you will bear with me as I toil through it.

In a certain classified ordnance application, two HE weapons are detonated, and it is a matter of concern whether the shock waves from the explosives arrive at a point between them simultaneously and before the occurrence of a particular event at the intermediate point.

Speaking generally, we have three events, each with its own distribution in time, and we want to find the probability associated with certain spacings and orders of occurrence of the events.

Calling the locations of the two weapons and the intermediate point A, C, and B, respectively, as illustrated in figure 2,

A______________B______________C

the problem is to determine the probability of arrival of shock waves from A and C at B simultaneously and before the occurrence of an event at B (called hereafter event B). Simultaneous is defined arbitrarily as within 100 micro-seconds. The expected times of the detonation and event B may be the same or different in some ordered manner.

For visualization purposes, it may be considered that the interaction effects of the two shock fronts are to be photographed using a Schlieren technique. The shock waves must meet within the brief angle covered

by the camera. The camera's film supply is limited; and event B is the start of film exposure. The 100-microsecond simultaneity period is the time within which the interaction is within the narrow range of the camera. This is not the real problem, but a hypothetical problem that I have invented, since the true problem is classified. I want to emphasize that it is the general method of solution, rather than the problem, that I want to present. To make the problem fit the solution, system failure must be thought of as coincidence of the shock waves at B but before the film has started running.

This paper will develop a procedure for determining the probability of system failure given the expected times of the detonation of the HE weapons, the expected time of event B, and the probability distribution of these times.

In the situation in which this problem arose, it was necessary to find a solution because the probability of the shock waves arriving at B simultaneously and before event B occurred was required to be very small, of the order of 0.001, while the variability of some of the proposed detonators and other components was of the same order as the shock wave travel time from A or C to B. It was necessary to find whether such variability could be tolerated, and, if not, how tight the dispersion had to be. In order to aid the required design decisions, it seemed desirable to get the results in the parameterized form of a plot of system sigma versus probability for selected shock wave travel times.

As a matter of personal preference, I looked for a desk calculator solution, which might later be programmed for computer.

In its simplest form, which I will discuss first, the problem has A and C equidistant from B, so that the shock wave travel times are equal, and all events will be expected to be absolutely simultaneous.

The problem requires finding, for every infinitessimal interval of time, the differential probability of system failure, which is the product of three probabilities -- the probability that A detonates within that infinitessimal interval; the probability that C detonates within 100 microseconds of the interval; and the probability that B has not occurred one shock wave travel time later than that interval. Integrating the product of these probabilities over all time gives us the total probability of

system failure. For the example used, the probability distributions of the event times were all taken as normal and the standard deviations as equal for A and C; but these assumptions are not necessary to apply the general method of solution.

Now I will put the problem in more general terms. In any infinitesimal period of time, dt, at a time t', the probability of system failure is the compound probability that event A has happened a constant, predictable time earlier than t', that event C has happened within a stated small interval about a constant, predictable time earlier than t', and that event B has not yet happened at time t'. The constant, predictable times are the shock wave travel times from A and C to B, which may or may not be equal in the general case; and the stated small interval is the simultaneity period, which must be small enough relative to the travel times that events occurring within it may be considered simultaneous.

Integrating this compound probability over all time yields the total probability of system failure.

As illustrated in figure 3, we will call the times of occurrence of events, A, B, and C, t_A, t_B, and t_C, and the probability distributions of these events, $P(t_A \geq t)$, $P(t_B \geq t)$, and $P(t_C \geq t)$. The shock wave travel time depends largely on the travel distance and on the amount of explosive involved, and is considered to be constant. We will call the travel times from A to B and from C to B, t_{AB} and t_{CB}. The short period within which shock wave arrival is considered simultaneous, we will call Δ.

Starting with the probability distributions of the times of events A, B, and C, each of which has somewhat of the appearance of the top curve of figure 4, we proceed as follows to find the probability of system failure.

For convenience in notation and in thinking about the problem, we will tie our general time frame to the time frame of event A. This poses no difficulty since the expected times of events A, B, and C are known; and we would, in any event, have used one of these fixed times as the origin of the general time system.

For a system failure to occur at time t', then, $t' = t_A + t_{AB}$. Also, for the necessary simultaneity, t_C must occur within the period Δ and later than t_A by the difference between their traveltimes to B; or, in mathematical language, as it is written on the figure 3. (Of course, it is an algebraic "later" and, if $t_{AB}-t_{CB}$ is negative in sign, t_C must occur earlier in time than t_A for simultaneity of shock wave arrival at B.)

Finally, for system failure, when the simultaneous shock wave arrives at B at time t', event B must not have occurred. Therefore, $t_B \geq t_A + t_{AB}$.

The probability that event A occurs within any differential period of time, dt, is $d\,P\,(t_A \geq t)$. This differential probability must be multiplied by the probability that event C occurs in an interval, Δ, $\overline{t_{AB}-t_{CB}}$ later than dt, $P\,(t_A + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t_A + \overline{t_{AB} - t_{CB}} + \Delta)$.

The product must further be multiplied by the probability that event B occurs after t', $P\,(t_B \geq t_A + t_{AB})$. Integrating this final product over all $P\,(t_A \geq t)$ is equivalent to integrating over all t, since $P\,(t_A \geq t)$ is a single valued function of t.

The probability that C occurs within the simultaneity interval of any time t is obtained from the probability distribution of the times of event C. In the case of normal distributions, this is easy to do.

The curve of this function versus t has the general form of the bottom curve of figure 4. Then the probability distribution, $P\,(t_B \geq t)$ versus t is modified to $P\,(t_B - t_{AB} \geq t)$ versus t.

These two functions of t are multiplied together to get $P\,(t_B - t_{AB} \geq t)$. $P\,(t + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t + \overline{t_{AB} - t_{CB}} + \Delta)$ versus t.

Since $P\,(t_A \geq t)$ is a single valued function of t, values of this probability can be substitued for values of t to get a plot of $P\,(t_B - t_{AB} \geq t)$

times $P(t + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t + \overline{t_{AB} - t_{CB}} + \Delta)$ versus $P(t_A \geq t)$ (as in figure 5). The area under this last curve is;

$$\int_0^1 P(t_B \geq t + t_{AB})\, P(t + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t + \overline{t_{AB} - t_{CB}} + \Delta)\, dP(t_A \geq t),$$

and this is the probability of a system failure, P_{SF}. The important attribute of this method of solution is that this area may readily be evaluated, without constructing the curves, by use of Simpson's Rule.

If a Simpson Rule division of the area into ten parts gives us enough accuracy (as it well may, depending on how accurately we know the shapes of the distributions involved), we need only find eleven values of t corresponding to $P(t_A \geq t)$ values of 0, 0.1, 0.2, etc., to 1.0; and two of these times are plus and minus infinity. At these extreme times, the simultaneity probabilities are zero. For the intermediate times, the simultaneity probability may easily be looked up in any well-detailed table of areas under the normal curve for the normal distributions assumed in our example.

A sample computation has been shown in figure 1. For any computation using a 10-part Simpson Rule integration, the first two columns will be the same. To get the simultaneity probabilities, it is observed that $\Delta = 0.01 \cdot \sigma_A$; so that for the second time point the probability looked up is that of being between 1.292 and 1.272 standard deviations away from the mean. For the column of $t + t_{AB}$, it is noted that $t_{AB} = 1\sigma_B$ and that $\sigma_A = 2\sigma_B$. Then since $\bar{t}_A = \bar{t}_B$, $\bar{t}_A + k\sigma_A = \bar{t}_B + (2k + 1)\sigma_B$. Having found these t_B equivalents, the table look-up is easy. The next column is the product of the third and fifth columns. These products are multiplied by the Simpson Rule factors, 1, 4, 2, 4, 2, etc.; and the sum is multiplied by the class interval of 0.1 and divided by 3 to get the value of the integral (see figure 4), which is the answer sought. Repetition enables the computation to be performed in about 20 minutes.

Taking expected times as equal, and at least two of the sigmas as equal, allows results to be plotted as in figure 6. Other conditions are not much harder to compute, but the results are harder to present.

Many problems besides this fictionalized and hypothetical one are susceptible to this technique of solution. The method is one which, with some familiarity, enables an engineer or mathematician to solve problems involving probability at his desk before or without submitting them for computer solution. And, of course, it is useful for those who have no access to a computer.

SAMPLE COMPUTATION

$\sigma_A = \sigma_C = 10$ M SEC [$= 2\,\sigma_B$] $\qquad \sigma_B = 5$ M SEC $\qquad T_{AB} = T_{CB} = 5$ M SEC [$= \sigma_B$] $\qquad \Delta = 0.1$ M SEC [$= .01\,\sigma_A$]

$P\ T_A \geq T$	T	P_{SIMUL}	$T+T_{AB}$	$P[T_B \geq T+T_{AB}]$	$P_{SIMUL}\,P[T_B \geq \cdot\cdot]$	SIMPSON RULE FACTORS
0	$\overline{X_A} + \infty$	0	$\overline{X_B} + \infty$	0	0	1
.1	$\overline{X_A} + 1.282\,\sigma_A$	.0035	$\overline{X_B} + 3.564\,\sigma_B$	.0002	.00000	4
.2	$\overline{X_A} + .842\,\sigma_A$	.0056	$\overline{X_B} + 2.684\,\sigma_B$	.0036	.00002	2
.3	$\overline{X_A} + .524\,\sigma_A$	.0070	$\overline{X_B} + 2.048\,\sigma_B$	.0203	.00014	4
.4	$\overline{X_A} + .253\,\sigma_A$	.0077	$\overline{X_B} + 1.506\,\sigma_B$	.0660	.00051	2
.5	$\overline{X_A} + 0$	.0080	$\overline{X_B} + 1.000\,\sigma_B$	.1587	.00127	4
.6	$\overline{X_A} - .253\,\sigma_A$	.0077	$\overline{X_B} + .494\,\sigma_B$	.3107	.00239	2
.7	$\overline{X_A} - .524\,\sigma_A$	.0070	$\overline{X_B} - .048\,\sigma_B$	.5191	.00363	4
.8	$\overline{X_A} - .842\,\sigma_A$	.0056	$\overline{X_B} - .684\,\sigma_B$	.7530	.00422	2
.9	$\overline{X_A} - 1.282\,\sigma_A$	.0035	$\overline{X_B} - 1.564\,\sigma_B$	.9411	.00329	4
1.0	$\overline{X_A} - \infty$	0	$\overline{X_B} - \infty$	1.	0	1
					$\Sigma = .04770$	

SYSTEM FAILURE PROBABILITY $= \dfrac{0.1\,\Sigma}{3} = .0016$

FIGURE 1

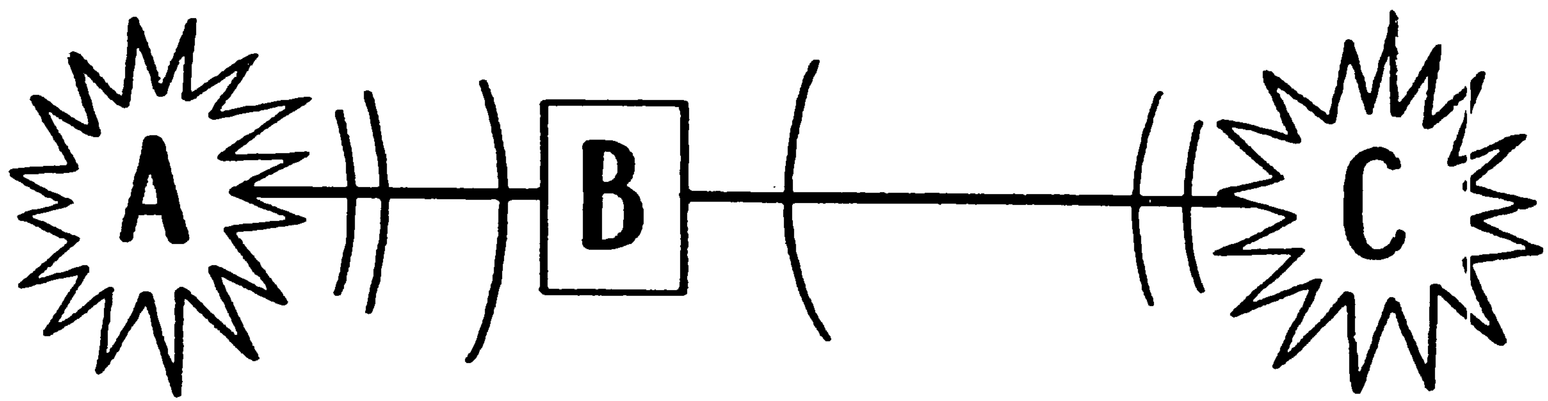


FIGURE 2

Event Times	t_A	t_B	t_C
Probability Distribution	$P(t_A \geq t)$	$P(t_B \geq t)$	$P(t_C \geq t)$
Shock Wave Travel Times	t_{AB}	t_{CB}	
Simultaneity Period	Δ		

Simult. of A and B $t_A + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t_A + \overline{t_{AB} - t_{CB}} + \Delta$

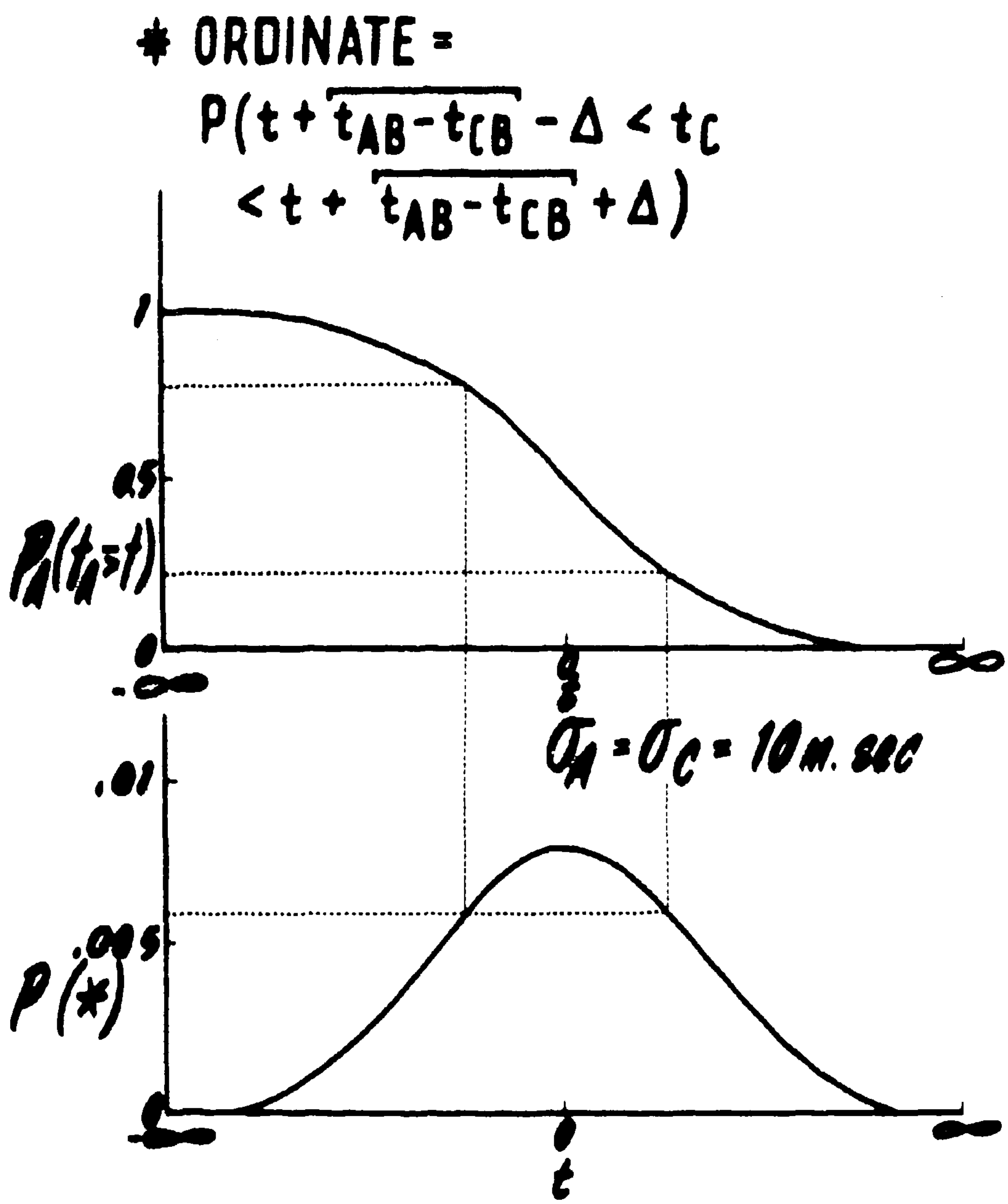
Simult. Shock Wave Arrival

At B Before Event B $t_B \geq t_A + t_{AB}$

Total System Failure Probability

$$\int_0^1 P(t_B \geq t + t_{AB})\, P(\textit{Simult. of A and B})\, dP(t_A \geq t)$$

FIGURE 3



* ORDINATE =
P(t + tAB - tCB - Δ < tC < t + tAB - tCB + Δ)
1
0.5
PA(tA=t)
0
-∞
0
∞
σA = σC = 10 m. sec
.01
.005
P(*)
0
-∞
0
∞
t

FIGURE 4

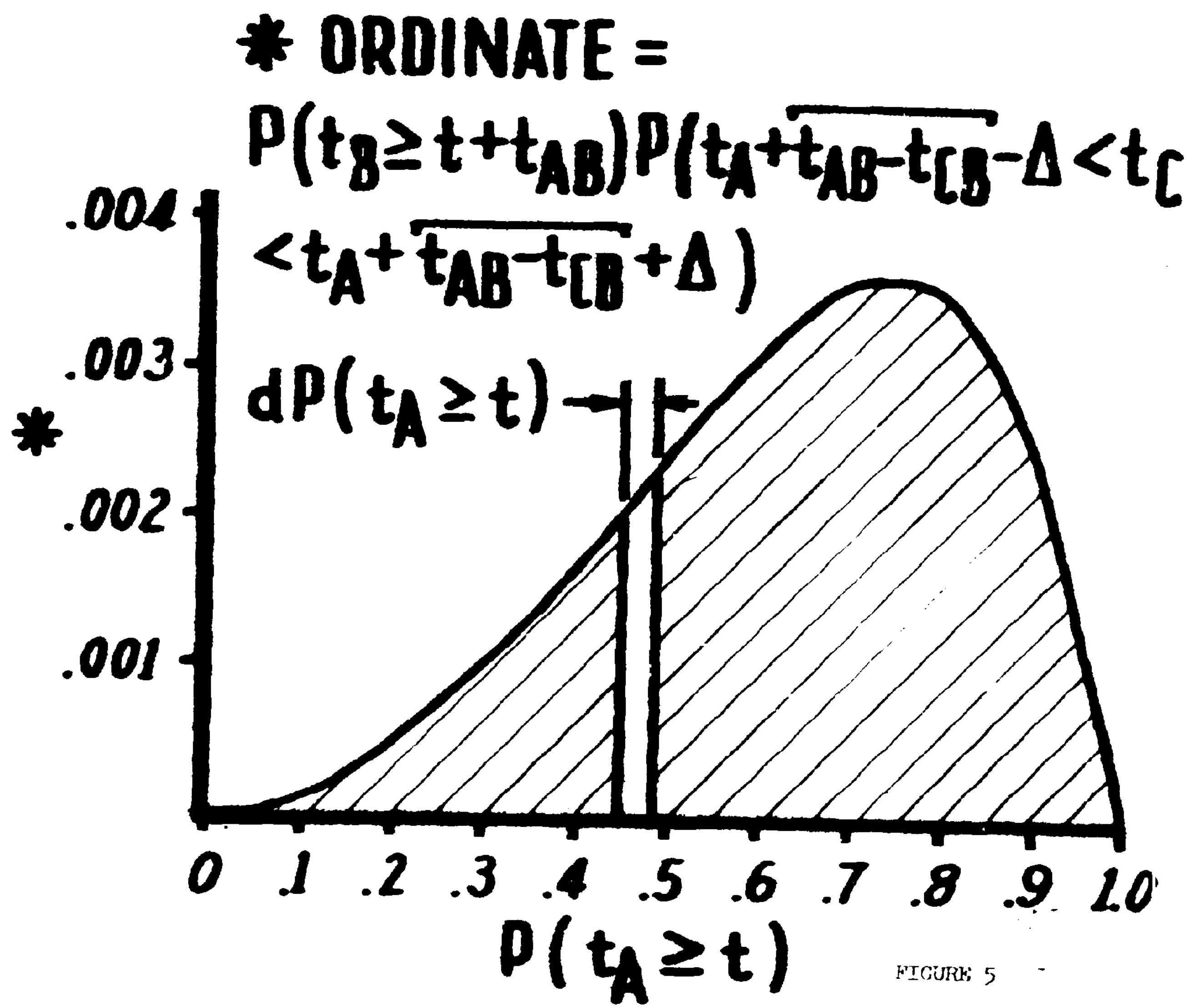
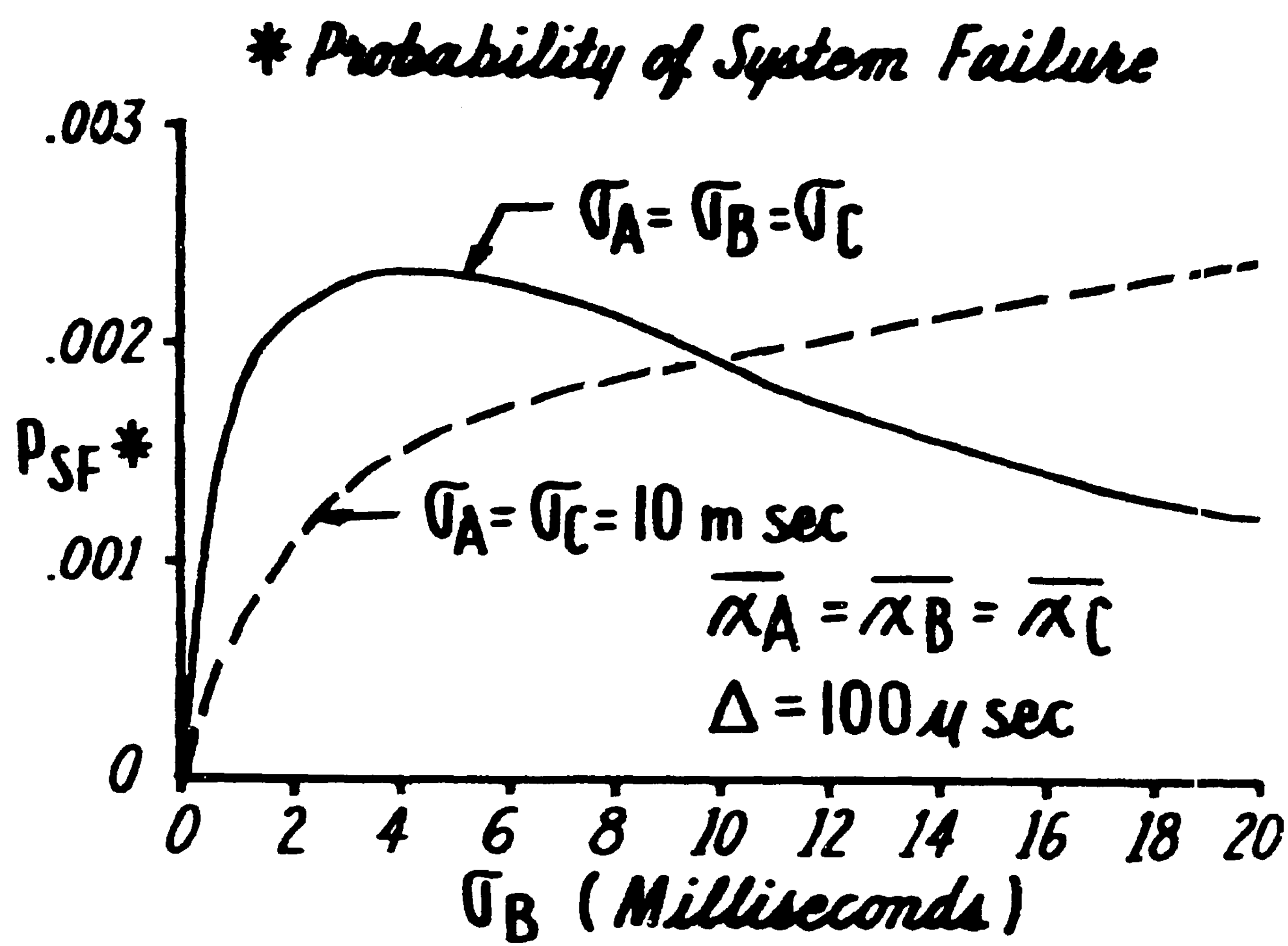

* ORDINATE =
P(tB ≥ t + tAB) P(tA + tAB − tCB − Δ < tC < tA + tAB − tCB + Δ)
dP(tA ≥ t)
*
.004
.003
.002
.001
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
P(tA ≥ t)

FIGURE 5



* Probability of System Failure
.003
.002
P_{SF} *
.001
0
$\sigma_A = \sigma_B = \sigma_C$
$\sigma_A = \sigma_C = 10$ m sec
$\overline{\lambda_A} = \overline{\lambda_B} = \overline{\lambda_C}$
$\Delta = 100\,\mu$ sec
0 2 4 6 8 10 12 14 16 18 20
σ_B (Milliseconds)

FIGURE 6

A DATA COLLECTION PROCEDURE FOR ASSESSING NEURO-MOTOR PERFORMANCE IN THE PRESENCE OF MISSILE WOUNDS

William H. Kirby, Jr., William Kokinakis,
Larry M. Sturdivan, and William P. Johnson
U. S. Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

INTRODUCTION. While medical clinicians diagnose, treat, and judge the sequels of injury, it is of interest to others such as those engaged in man-task or man-machine system studies to consider the effects of injury from additional points of view. Those concerned with the medical problems are naturally interested in procedures and optimal treatments which prevent or at least minimize the consequences of injury. Those concerned with man-machine system performance problems are interested, in addition, in the ability of the injured or otherwise stressed individual to perform given tasks. Common to both the clinician and the man-machine researcher interested in injury is a need for a better understanding of the mechanisms and responses associated with traumatic pathological dynamics.

We are in the process of developing a methodology for describing and assessing anatomical and physiological pathologies associated with missile wounds. Hopefully, we will be able to express these in terms of a set of affectors and/or effectors such as those found in the nerve-muscle or neuro-motor structure. The reason for this approach is that they may serve as a common denominator for describing injury as well as for describing the task or machine operation requirement. In this presentation we will limit our attention to wounds caused by missiles. The type and amount of data to be collected will probably be influenced by the number of accident cases that enter hospitals which are accessible for study.

This is an interdisciplinary problem area in which clinicians, engineers, mathematicians and statisticians should meet. It is usually the case that such multi-discipline representatives are faced early in the process with certain communication problems. For example, a function to the clinician has one meaning, but to the mathematician it has quite another. A medical researcher may apply a Chi-Square test to a set of data in which the statistician may insist that the application is invalid due to the fact that the data do not conform to a normal distribution. A surgeon

may be entirely satisfied that the maximal strength of a grasp is equal in both hands of a patient as determined through a hand squeezing process whereas the engineer is satisfied only if such an assessment is in quantitative terms such as a pressure-time history. Clinicians can really get confused whenthey attempt to understand differences between mathematicians and statisticians.

DISCUSSION. Our first approach considers the body as a system composed of a set of clinical subsystems coordinated to maintain life and control human performance. These clinical subsystems will initially be divided for convenience into a primary and secondary group. The primary group will include the neurological, cardiovascular, respiratory, skeletal, and muscular sybsystems. The secondary group will consist of the gastrointestinal, genitourinary, and endocrine subsystems. While we intend to collect some data associated with the secondary group, initially only the primary group will be considered in detail.

We attempt to describe performance in terms of a simplified set of afferent and efferent (input-output) factors shown in Figure 1. For the present we intend only to recognize the presence or absence of the afferent (input) factors - vision and hearing, and the efferent (output) factor - voice. Essentially then we have reduced our performance descriptors to the first six listed in Figure 1. Actually these descriptors are regional subdivisions of the human body and they will be represented by the (motor) muscles which are located in the respective regions. The neurological or muscle activator network is distributed over these regions and no controlled human actions occur without its activation. It was therefore natural to choose these motor factors as a common denominator to which all performance phenomena and subsystem changes may be related.

Performance Factors

Performance Symbol	Efferent (Output) Factors
e_1	Right upper limb
e_2	Left upper limb
e_3	Right lower limb
e_4	Left lower limb
e_5	Head and Neck
e_6	Trunk
e_7	Verbal communication
	Afferent (Input) Factors
e_8	Vision
e_9	Hearing

FIGURE 1

Using the above rationale we are interested in collecting data from accidental wound cases in order to describe, classify, and relate important missile characteristics, clinical subsystem injuries, and neuro-motor performance phenomena. Concurrently a more comprehensive study of this type of problem is being considered but which is beyond the scope of this presentation.(1)

The Neuro-Motor or Effector Logic. The "terminal" body tissue or structure directly responsible for physical movements as indicated above is muscle. Inasmuch as muscles are innervated by specific peripheral nerves, associated nerves and muscles have become known as "neuro-motor units." Fortunately the nerve-muscle anatomical distribution system has been well established by anatomists in the past.

In order to demonstrate this logic attention is directed to Figure 2 which is a matrix showing the muscles and their actions in the upper limb. This matrix could represent either of the effectors, e_1 or e_2

since they are symmetrical. The numbers along the abscissa are subscripts of the letter "A" in which each subscript represents a specific anatomical action as described in Appendix A. The numbers along the ordinate are subscripts of the letter "M" in which each subscript represents a specific muscle also described in Appendix A. The muscles (m_i, $i=1,2,\ldots 61$) are arranged in a manner such that the lower numbers represent muscles in the shoulder and in ascending order represent muscles in the arm, forearm, and hand.

The distribution of nerves and their contained fibers which innervate the skeletal muscles is unique. For instance, the large number of nerve fibers which originate from a given source such as a particular spinal cord segment, are dispersed into a multiplicity of branches. As if to provide maximum reliability, many fibers from the same source reach a given muscle by different pathways. On the other hand nerve fibers from the same spinal cord level are known to innervate different muscles. While the nerve pathways are not demonstrated here, some idea of the nerve fiber distribution may be obtained from the left side of Figure 3. The C_i ($i = 1, 2, \ldots 8$), T_1, and T_2 represent nerve roots (large groups of fibers) which emerge from the designated spinal cord levels. These are identified in Appendix A. The letter, C, refers to the cervical or neck region and the subscripts refer to the specific locations out of which the bundles of nerve fibers flow. The letter, T, refers to the thoracic or chest region. Only the first two thoracic nerve bundles, T_1 and T_2, are included in Figure 3.

It is interesting to observe from this matrix that a given muscle is innervated by nerve fibers from more than one source. Note, for example, that M_4 (pectoralis major) is innervated by nerve fibers derived from several spinal cord segments, namely, C_5, C_6, C_7, C_8, and T_1. An important implication in conjunction with the previous comments on reliability is that if the spinal cord were severely injured at the level of C_8, muscle, M_4, would not become completely paralyzed inasmuch as it would still receive considerable innervation from fibers above the site of injury namely, C_5, C_6, and C_7. It is also interesting to observe the number of muscles in the shoulder region innervated by a given nerve root such as C_5. They include M_2, M_3, M_4, M_7, M_{10}, M_{11}, M_{12}, M_{13}, M_{15}, M_{16}, M_{17}, M_{18}, and M_{19}.

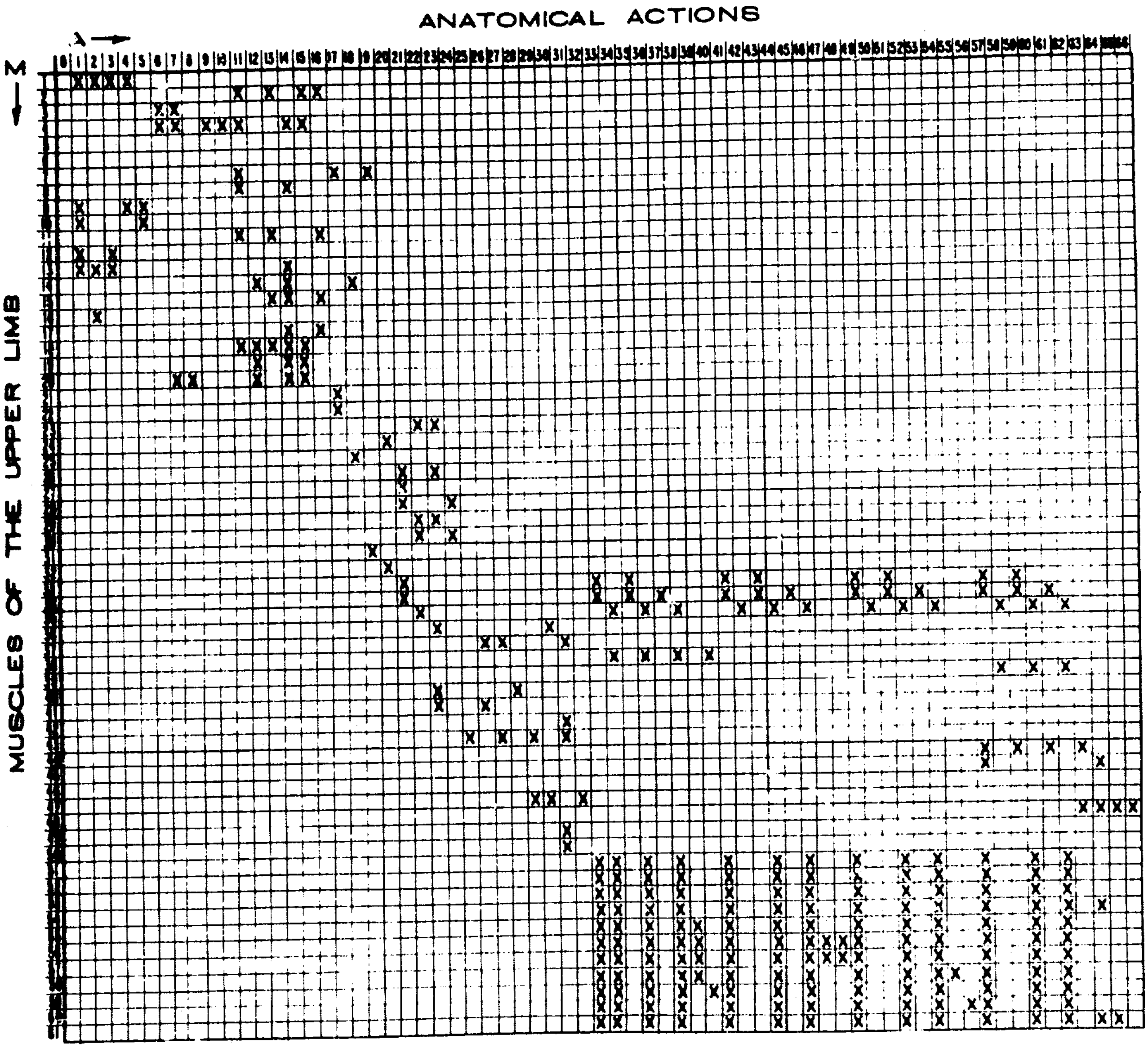


TABLE I

ANATOMICAL ACTIONS VS. MUSCLES OF THE UPPER LIMB

$A_{0,1,2}$ -------66 REPRESENT ANATOMICAL ACTIONS

$M_{1,2}$ --------61 REPRESENT UPPER LIMB MUSCLES

FIGURE 2

INNERVATING NERVE FIBERS — ANATOMICAL ACTIONS

A→

T_2	T_1	C_8	C_7	C_6	C_5	C_4	C_3	C_2	C_1	M ↓	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
						X	X			1		X	X	X	X																
				X	X					2												X		X		X	X				
				X	X					3							X	X													
	X	X	X	X	X					4							X	X		X	X	X			X	X					
							X	X		5																					
							X	X	X	6																					
				X	X					7												X						X		X	
			X	X						8												X			X						
	X	X								9		X			X	X															
			X	X	X					10		X				X															
					X					11												X		X			X				
					X	X	X			12		X		X																	
					X					13		X	X	X											X						
		X	X							14													X		X				X		
				X	X					15														X	X		X				
					X					16			X																		
					X					17															X		X				
				X	X					18												X	X	X	X	X					
				X	X					19													X		X	X					
		X	X	X						20									X				X		X	X					

MUSCLES OF THE SHOULDER

TABLE II

NERVE-MUSCLE-ACTION RELATIONS OF THE SHOULDER

A_0 ,1,2——20 REPRESENT ANATOMICAL ACTIONS

M_1 ,2 ······ 20 REPRESENT SHOULDER MUSCLES

C1, 2, ··· C_8, T_1, T_2 REPRESENT INNERVATING SPINAL NERVES

FIGURE 3

This multiplicity of overlap or redundancy is unique for minimizing the effects of injury. Quantitative force time relationships have not been effectively established at the muscle level to allow us to assign weighting factors to the contributions from a muscle to an associated anatomical action.

Single muscles and their respective anatomical actions combine to characterize regional and joint actions. In the shoulder region, for instance, we think first in terms of the muscles associated with the specific anatomical actions such as flexion, extension, abduction, and adduction. In continuing the generalization process in this region we next consider the shoulder motions of rotation and circumflexion which are derived from the same muscles acting in different combinations and sequences. In applying the same notions to the hand we begin by considering the single muscles associated with finger flexion, extension, abduction, adduction and opposition and then combining these in ways to account for generalized processes such as grasping, holding, and releasing. They, of course, are associated with even more complicated processes associated with performing tasks such as using a screw driver or turning a door knob.

In brief we hope to be able to associate the biomechanical functions with the natural effectors or neuro-motor factors which are responsible for them. One may proceed in man-task study problems in either direction, i.e., he may begin with the knowledge of basic muscular functions and move up the scale to gross movements or he may start with a study of the man-task process in the hope of first identifying useful gross movements and work down to the scale to muscle functions.

A combined mechanical and anatomical orientation appears to have some unique advantages for describing man-machine interaction. For example, we believe that by considering the upper limb as a flexible multi-jointed cantilever with a unique prehensile of grasping device located at the free end, one can develop useful methods for describing physical and physiological factors in relation to man-machine interactions in ways which yield to simplification and measurement.

In our first approach to upper limb biomechanical measurements we are using these anatomical-mechanical notions, i.e., muscle groups associated with hand actions, joint actions, multi-joint actions, liner

actions, and combinations of these. Initially we propose to measure only a limited number of these biomechanical or effector functions. For the upper limb we have developed some instrumentation to measure and record force-time histories associated with hand grasping, and flexion and extension actions about the wrist and elbow joints. A brief explanation of this instrumentation is given in Appendix B.

We extend this rationale to include the opposite upper limb, the two lower limbs (considering them also as multi-jointed cantilevers but in terms of their natural anatomical-mechanical functions of weight-bearing and ambulation), and the other effectors (head and neck, and trunk). We believe that this approach will result in useful descriptors for man-task/machine interactions in a manner suitable for describing and assessing changes in performance due to disability regardless of cause.

CLINICAL SUBSYSTEMS. One of the reasons this problem is of interest is that it requires investigations not made in the past. For example, it is known that damage to the cardiovascular subsystem in terms of blood loss is likely to be fatal if the value exceeds approximately 1600 cc to 1700 cc within a short period assuming no replacement. While this is an important upper bound, the effect on one's ability to perform due to hemorrhage of lower orders has not to our knowledge been studied. Hence it is of primary interest for us to collect hospital data on patients who may suffer various degrees of blood loss and to measure the effects on several representative effectors.

Damage to the respiratory subsystem may be assessed in terms of the degree of pneumothorax, rate of O_2 - CO_2 exchange, or, perhaps, in terms of respiration rate and depth. The chosen effectors would be measured at approximately the same time that the physiological measures are taken.

Descriptions of levels of damage may be more difficult for the neurological, muscular, and skeletal subsystems. Presently we are only considering two levels of damage for any substructure (a muscle, a nerve, or a bone), namely, none or complete. For the present we do not expect to make special studies on the gastro-intestinal, genitourinary and endocrine subsystems other than to observe the routine hospital events associated with them. Their respective blood losses are to be considered, however, but viewed as cardiovascular subsystem deficits.

Additional measurements which are not always routine may be added if the results of some of the present research being done elsewhere on traumatic injuries indicates it. For example, one of us is engaged in human shock research in which certain relations between clotting time and levels of shock have been generated for 30-odd humans who were in shock due to blood loss.(3) Interesting observations of adrenal function in combat and wounded soldiers have been made to some extent by others. (4) (5) These suggest possibilities for associating "stress" levels and performance.

It is with these ideas in mind and the notion that feasible relations between the effectors and the body's clinical subsystems do exist that we wish to construct useful data collecting procedures. Hopefully, early insight following some of this data collection will allow us to get some ideas concerning these relations. Since we expect such potential relationships to change with time as a result of injury, we believe that data collection will have to be made at various time periods throughout the clinical course.

Having set up these ideas as guidelines for the data collection process, we must consider some of the practical aspects of the problem. It is important to review the clinical procedures used in evaluating and treating wound cases in hospital accident rooms, operating rooms, and recovery wards in order to appraise the available and/or recorded data in terms of type and quantity. Another point of interest concerns the logic for selecting and running certain clinical tests and not others. There may be many cases in which certain useful clinical data could be made available for specific purposes such as ours but which may not be sought ordinarily by a clinician inasmuch as these data do not in his judgment add any useful information for his purposes. It is also important to be sure that the acquisition of data from a distressed patient does not interfere with his well-being.

Medical Records and Clinical Data. Hospital medical records reflect traditional procedures for recording information and events associated with the professional care and treatment of patients. While time and space preclude any extensive discussion of the meaningfulness of the comprehensive clinical and laboratory data as interpreted by physicians and/or other interested discipline representatives, a few observations are presented.

Clinicians evaluate patients, their care and theraphy, according to the description and history of the presenting complaint as well as other pertinent past patient (and family) history, physical examination, laboratory test results, and progress evaluations. In general clinical information is classified either as subjective or objective information. Subjective information is associated with what the patient or others tell to the examiner. Examples of this would include the patient's interpretation of local or general muscular weakness, walking difficulty, ------ dragging toe of shoe, stumbling or falling, sphinteric disturbances (inability to hold urine), changes in local or general sensation ---- fixed or radiating pain, temperature, tactile discrimination, deep sensation (muscle, bone, vibratory sense), and abnormal sweating. Objective information concerns what the examiner learns from his own observations such as range of limb movement, contractures, diminished size, strength of muscles against resistance, tremors, etc. Thus, it can be seen that except for laboratory tests and some clinical items such as blood pressure, pulse rate, and the electrocardiogram, quantitative measures are minimal in the traditional records.

Records of emergency cases are often initiated in the hospital's accident room and accompany the patient throughout his hospital course. A form of time history of his care, diagnosis, treatment, and progress is recorded for permanent file. We will consider a few examples of hospital records.

An accident room record is shown in Figure 4. This 21 year old male's record shows very little information. This is not unusual in the typical busy accident room. The only history and physical data recorded in this case are blood pressure and pulse. The immediate treatments and the results of laboratory blood tests were recorded. It is highly likely that additional observations of blood pressure and pulse were made with the passage of time but not entered in the record. A certain amount of physical examination was probably performed and not recorded. While the admission time and date were recorded, a detailed history of the presenting complaint is not shown. Such information might have included approximate time of the shooting incident, estimates of external blood loss, and conscious or unconscious behavior of the patient until arrival at the hospital, and ballistic factors, e.g., type and calibre of gun, distance between firer and patient, and angle of target to firer.

UNIVERSITY HOSPITAL
CENTRAL CITY

DATE 9-24-62 ARRIVAL 7:40 P

DEPARTURE A.P.M.

NAME Abbott Robert AGE 21 SEX M COLOR C BIRTHDAY RELIGION

ADDRESS 1022 King Rd PHONE

NEAREST RELATIVE ADDRESS

TIME OF ACCIDENT A P M AUTO COMP. B.C. M.C. OTHER

EMPLOYER & ADDRESS

INSURED BY COMPANY PHYSICIAN

ARRIVED VIA Ambulance E.R. PHYSICIAN Dr. Jones

COMPLAINT 2 Gunshot wounds One in chest and one in back

DATE & TIME OF ONSET OF ILLNESS OR INJURY Store - Pratt + 10th St. - Amsterdam Co.

ORDERS	PHYSICIANS RECORD Police officer Brown
On adm	BP 130/86 P-100
1000 cc Dextrose 5%	
in Saline	
500 cc Dextran	
Plasma to	
follow (300 cc)	
Tetanus toxoid 1cc IM	
✓Hb 14.5	
✓Hct 42	
✓WBC 9,600	
✓T-match (6 units)	
✓T+Hold	
✓BUN	
✓CO_2	
✓Sugar	

FIGURE 4

DISPOSITION DR'S SIGNATURE M.D.

Another accident room record is shown in Figure 5. This 15 year old male with multiple gun shot wounds has much more information on the physical examination than the previous case. Another example is shown in Figure 6. One must judge that in this latter case the patient was moved without delay to a ward bed for further control and treatment.

In examining a number of similar records, we believe that information which is useful for the clinician is, for our purposes, both inexact and insufficient. It is not enough for us to know that "the neurological and extremities are negative." It is our opinion that we need to have some measured data associated with the neuro-motor activity of the effectors. For example, the prehensile or grasping function of an upper limb although not injured may be weakened due to a severe blood loss in an artery located in another part of the body. Neurologists and neurosurgeons have a variety of clinical tests which they use based on certain known relationships in neuro-anatomy and physiology. As a matter of interest, it has been stated that in no other branch of medicine is it possible to build up a clinical picture so exact as to localization of pathological conditions.[2] However, if such tests are not performed and/or the results not recorded for patients of interest to us, possible relations between the effectors and clinical subsystem changes are not attainable.

It may be noted also that many classical neuro-motor tests are not strictly quantitative as evidence by such observational descriptions as "partial paralysis" or "some loss of sensation." Conclusions drawn from a variety of tests may vary considerable as they depend on the judgment of different examiners and the impressions expressed by patients.

In the ward the situation is quite different from that found in the accident room. Here the patient is in an environment in which physicians, nurses, and technicians are available to care and control the patient under more favorable conditions. Records are kept in much more detail. The frequency of observations and the application of procedures and treatments are, of course, greater when the patient is in serious condition. Except for emergency procedures, these observations and treatments for patients under intensive care are usually performed about every one of two hours. It takes this long for many therapies to take effect.

UNIVERSITY HOSPITAL
CENTRAL CITY

659

DATE 1/29/63 ARRIVAL 3°° PM
DEPARTURE A.P.M.

NAME Jones, William AGE 15 SEX M COLOR N BIRTHDAY RELIGION
ADDRESS 102 N. Charles St Jamestown PHONE None
NEAREST RELATIVE father – Joe ADDRESS Same
TIME OF ACCIDENT A.P.M. AUTO COMP B.C M.C OTHER
EMPLOYER & ADDRESS
INSURED BY COMPANY PHYSICIAN
ARRIVED VIA Jamestown Rescue Amb 42 E.R. PHYSICIAN
COMPLAINT Gunshot wound Ⓡ arm & hand
DATE & TIME OF ONSET OF ILLNESS OR INJURY

ORDERS	92–20 150/90 PHYSICIANS RECORD
Hct, Hgb, WBC,	
Type & Cross	Pt brought into AR c̄ bullet wounds
4 units	Drawn
Bun, sugar	Inspection: Three openings – two in Ⓡ upper
500 cc D/Ringer	arm – in biceps muscle
	Third in Ⓡ chest wall about level of
	T5 and axillary line
#14	
Nasogastric tube	Sensorium Clear
inserted	Vital Signs BP 140/80 pulse: 80 resp: 16
	Chest
	Or Sounds thought to be decreased in Ⓡ hemithorax
	with dullness to percussion over both lung fields
	to about level of T6
	No Subcut. Emphysema
	Cor Sinus Rhythm No Ⓜ
Tetanus 1cc	Abd B Sds present
	Tenderness in epigastrium
	No liver enlargement
	Extrem: palpable pulse in Ⓡ arm
	With no sensory deficit
	Imp: Bullet wound
	R/O Ⓡ hemopneumothorax

FIGURE 5

DISPOSITION DR'S SIGNATURE [signature]

UNIVERSITY HOSPITAL
CENTRAL CITY

601

DATE 11-12-62 ARRIVAL 6 A.P.M.
DEPARTURE A.P.M.

NAME Brown, Walter AGE 52 SEX M COLOR W BIRTHDAY 9/4/10 RELIGION

ADDRESS 1505 N. Bond St. PHONE

NEAREST RELATIVE Mabel Brown (wife) ADDRESS same

TIME OF ACCIDENT A.P.M. AUTO COMP B.C. M.C. OTHER

EMPLOYER & ADDRESS

INSURED BY COMPANY PHYSICIAN

ARRIVED VIA Pittsville Vol. Amb. E R PHYSICIAN Dr. Jackson

COMPLAINT was robbed + shot - left chest @ intersection of Rts. 9 + 49

DATE & TIME OF ONSET OF ILLNESS OR INJURY 2 miles west of Pittsville

ORDERS	PHYSICIANS RECORD
	Gunshot wounds left chest with pulmonary hemorrhage
	X-ray = fragmented bullet in substernal chest wall on left with parenchymal hemorrhage mid lung field. No Pneumothorax
	Cauterize shallow wound
	admit for observation
	J. J. Johnson

FIGURE 6

DISPOSITION DR.'S SIGNATURE M.D.

Whenever a patient undergoes a special procedure, operation, or, if an autopsy is performed, a detailed account of the events is made which becomes a part of the permanent medical record. The first detailed physical examination is often not performed until the patient arrives in the ward following evaluation and treatment in the accident room. Additional laboratory tests and special diagnostic procedures are usually initiated after admission to the ward. A considerable amount of clinical information is recorded which can be useful for our purposes. Large gaps invariably exist, however, due to reasons mentioned previously as well as the thoroughness of the work and recording of interns and residents. Patient care is naturally oriented toward healing and care processes. This is to say that attention is on the progress of the patient's subsystems and behavior as a whole while he is in a resting state. He is not considered as a component in a man-machine system.

A Given Wound Patient Record. In order to get an idea of some of the events which may take place in an accident room and a ward, a medical record of a patient admitted with a bullet would is briefly discussed.

Accident Room.

1st. Day: A 56 year old man was shot in the chest with a 32 Calibre pistol "at close range." He was taken to a hospital and arrived at 5:36 a.m. on the day of the accident.

On admission to the accident room, the following were studied or measured immediately and recorded:

Blood Pressure
Pulse rate
Heart sounds
Breath sounds
Hemoglobin
Chest X-ray

Treatment was also ordered and initiated immediately. It consisted of a fluid replacement program according to the following sequence:

Circulatory expander started (500 cc)
Whole blood (500 cc)
Saline (500 cc)
Dextrose in water (1000 cc)

The patient was also sedated and prevented from taking any fluid or food by mouth (as a precaution in case of the need for operation).

The remaining accident room events which occurred according to the medical record were:

Follow-up blood pressure, pulse rate, and respiration checks at 8:00 am, 8:25 am, 9:00 am, 10:00 am, 12:00 noon, 4:00 pm and 6:00 pm.

The patient was admitted to the ward at 6:30 pm.

Ward The admitting intern made the first detailed investigation at 7:30 pm. It is shown as follows:

Intern's Admission Note: "At 4:00 am patient was shot by wife at close range with a .32 Calibre pistol."

Bullet entered left chest and escaped via left back.

Patient did not lose consciousness.

There was no chest pain, coughing increase, or hemoptysis (coughing of blood).

Physical Examination: Well developed, well nourished, alert, cooperative, no distress (two?) gun shot wounds, no powder burns.

Entrance: approx. 5 mm left anterior axillary line, 8th interspace

Exit: Post axillary line, 10th interspace

Chest: Expands well bilaterally
Lungs - left posterior: basilar dullness, about 3 cm above the base; tactile fremitus (a vibration imparted to the hand placed on the chest) somewhat impaired left posterior base.

Blood Pressure: 120/60 - no murmurs

Abdomen: Slightly distended
Generalized superficial tenderness

Neurological and Extremities - negative

Impressions: (1) gunshot wound, left chest
(2) left hemothorax
(3) rule out perforated bowel or spleen
(4) Cardiac arrythmia - bigemini (paired pulse beats)

Additional medications were given at 8:00 pm.

2nd. Day: Throughout the second day the following vital signs were observed every hour.

Blood pressure
Pulse rate
Respiration rate
Temperature

Nourishment was still given by intravenous fluids.

3rd. Day: Vital signs were observed every four hours instead of the hourly schedule of the previous day. A liquid diet was prescribed in place of the intravenous fluids.

After three more days the patient was placed on routine care. Except for a thoracentisis (extraction of fluid from the chest cavity via needle and syringe) on the 8th, hospital day (250 cc of bloody fluid was removed), the patient recovered and was discharged on the 12th. day.

Inasmuch as no operating room procedures were required, there was no opportunity to get a pathological description of the internal path of the bullet. In operative cases wound tract descriptions are usually available in varying amounts of detail. In many instances some estimates of hematoma (trapped blood) volume, degree of bone fracture, and other gross abnormalities are noted in the operating room reports. In some cases, missiles or missile fragments are removed whereas in others

the additional risk associated with the removal process is such that the fragment(s) are left in the body, X-ray studies, of course, are informative in such cases.

In this case the calibre of the gun was known and some indication of distance between gun and target was given. Occasionally additional information is given which is quite important such as an account of the patient's response following the wounding process (e.g., "the patient ran to the doorway and down the steps before fainting," "the patient screamed and fell to the floor unconscious," etc.). Estimates of blood loss, if noticed, and elapsed time between the shooting incident and arrival at the hospital accident room are of basic interest to us.

From a cursory review of medical records concerning missile wounds, one is impressed with the fact that the large majority of such cases are non-lethal. In survival records it is observed that patient handling, treatments and control vary considerably depending on the case. However, the general care patterns do not appear to be radically different. The type and amount of data recorded on the other hand varies widely. Lethal cases, of course, usually have an autopsy report providing one was permitted by next of kin or ordered by a medical examiner.

Remarks on Hospital Medical Records. In brief several points concerning information and medical records for patients entering the hospital with gunshot wounds are presented.

1. There is considerable variation in type and content of reported information. This is particularly so in the accident room portion of the record.

2. While it may be assumed that a multitude of observations on a patient's subsystems, appearance, manner of behavior, speech, etc., are made by clinicians which may influence diagnosis, treatment, and control decisions, such information is not usually so extensively recorded.

3. Patient information is difficult to handle and retrieve inasmuch as it is not organized according to subsystem variables over time. Some of the measures which are considered as standard include

vital signs such as temperature, blood pressure, pulse rate, and respiration rate. These are usually plotted on graph paper by the nurses. Clinical descriptions and observations are often matters of judgment which do not yield to measurement under present state-of-the-art.

4. Facts and observations pertaining to shooting incidents and especially anatomical, physiological, and psychological events which occur between the incident and arrival at the hospital are usually very brief when included.

5. What is or is not recorded probably depends on the amount of training, experience, and judgment of the respective hospital staffs and residents.

6. Usually no psychological or psychiatric examinations are made on patients suffering from missile trauma. However, this is not done as a routine on traumatized patients in general.

Proposed Data Collection. We have described what we wish to do, a method for going about it, the type of data we think we need as a first approximation, and the entent of its availability in the (emergency) hospital. It is apparent that our requirements call for more complete and more frequent clinical observations and measures in addition to the new information in support of our special interests. This information may be thought of as:

1. Ballistic information (Appendix C).

2. Postwounding behavioral information (Appendix D).

3. Pathological information in terms of

   a. Wound tract information (Appendix E).
   b. Clinical subsystem information (Appendix F).

4. Effector measurement information (Appendix G).

5. Special studies for

   a. Information on normals (Appendix H).
   b. Anthropometric information (Appendix I).

It is appreciated that there are many practical problems in organizing and administering an effort of this kind. Since wounded patients are classified as surgical patients it seems logical that the heads of the participating surgical divisions and their supporting resident and nursing staffs should be sufficiently interested in a program such as this. Modifications in the program of data collection are anticipated once sufficient feedback information is developed.

Hopefully, as mentioned in the early part of this discussion, useful relations may be forthcoming early in the process between missile parameters and subsystem changes and between subsystem changes and the (biomechanical output) effectors. Our mathematical and statistical colleagues tell us that this effector-wound relationship is a stochastic one for two reasons. First, there is a distribution of wounds within a given class or category and hence there is a random variation with a certain distribution function of the corresponding effector. Second, the effect of even identical wounds sustained by different individuals also possesses a certain distribution.

There are many obvious problems involved in analyzing the data to be collected. Unfortunately there are no known data previously collected from which normals or non-injured standards can be established. It is, of course, impossible to obtain pre-injury effector and subsystem "normals" for the hospital cases. Since we are primarily interested in males of military age, the choice of civilian accidents for data collection will be so restricted. However, useful information may be gathered for cases in which injured patients would return following complete recovery. Otherwise independent samples will have to be taken from normal personnel in order to obtain pre-injury distributions of the appropriate subsystem and effector parameters. There is also a need for anthropomorphic data. Therefore they are included under special studies (Appendix I).

<u>Sources of Error in the Procedures</u>. The following are some of the more apparent errors which we are told should be considered:

1. Errors in determining the group normal for each effector measure inasmuch as (a) subjects cannot be evaluated before wounding and (b) a statistical group normal must be established with which to compare the individual disability.

2. Error arising from individual deviation from the group normal.

3. Error arising from interaction between the effectors. These errors are minimized by attempting to isolate the effectors, but they cannot be totally eliminated.

4. Errors arising from variation in the measuring and assessing techniques of the medical and technical evaluators.

5. Errors in the mechanical devices used to quantitate the efficiencies and errors associated with positioning the derives on the subjects.

6. Clinical laboratory errors.

7. Multi-clinic errors especially where more than one hospital is chosen for data collection.

SUMMARY. In an initial effort to quantify changes in human performance due to missile injury, we propose to collect certain clinical information, associated missile ballistic factors, and measure a select group of neuro-muscular responses. Such responses are chosen as potential descriptors for man-machine performance problems. As mentioned, a man-machine system model for incapacitation evaluation is being developed.[1] The acquisition of detailed human clinical data is essentially restricted to that available in the emergency hospital. Hopefully, such information will permit the establishment of new and useful relations early in the process.

REFERENCES

1. A Two-Person Zero-Sum Game Application for Incapacitation Evaluation, W. H. Kirby, Jr., and C. Masaitis. Ballistic Research Laboratories (in process).

2. Practical Neurological Diagnosis, by R. G. Spuring. Charles C. Thomas - Publisher, January 1950.

3. The Effect of Hemorrhagic Shock on Clotting Time in Humans, Attar, A., Kirby, W. H. Jr., Masaitis, C., Cowley, R. A. (in process).

4. Studies of Adrenal Function in Combat and Wounded Soldiers, Howard, John M. et al. Annals of Surgery, March 1955.

5. The Hemostatic Response to Injury. A Study of the Korean Battle Casualty. Annals of Surgery, March 1955, Scott, Russell, Jr., Crosby, W. H.

APPENDIX A

I. Anatomical Actions

Subscript codes for anatomical actions of the upper limb are given only for the shoulder region.

Subscript Code	Anatomical Actions
A_0	Actions of the Parts other than upper limb
A_1	Rotation of scapula
A_2	Adduction of scapula
A_3	Raising of scapula
A_4	Lowering of scapula
A_5	Moving of scapula forward
A_6	Moving of shoulder forward
A_7	Lowering of shoulder
A_8	Drawing of shoulder backward
A_9	Raise shoulder
A_{10}	Adduct shoulder
A_{11}	Flexion of arm
A_{12}	Extension of arm
A_{13}	Abduction of arm
A_{14}	Adduction of arm
A_{15}	Medial rotation of arm
A_{16}	Lateral rotation of arm
A_{17}	Flexion of forearm
A_{18}	Extension of forearm
A_{19}	Supination of hand by forearm
A_{20}	Pronation of hand by forearm

APPENDIX A (cont'd)

II. Muscles of the Upper Limb

Subscript codes for the muscles of the upper limb are given only for the shoulder region.

Subscript Code	Name of Muscle
M_1	Deltoid
M_2	Trapezius
M_3	Subclavius
M_4	Pectoralis major
M_5	Sternocleidomastoid
M_6	Sternohyoid
M_7	Biceps
M_8	Coracobrachialis
M_9	Pectoralis minor
M_{10}	Serratus anterior
M_{11}	Supraspinatus
M_{12}	Levator scapula
M_{13}	Rhomboid minor
M_{14}	Triceps
M_{15}	Infraspinatus
M_{16}	Rhomboid major
M_{17}	Teres minor
M_{18}	Subscapularis
M_{19}	Teres major
M_{20}	Latissimus dorsi

APPENDIX A (cont'd)

III. Spinal Nerves of the Upper Limb

Subscript codes for the spinal nerves which innervate the upper limb muscles are given in association with the shoulder region only.

Subscript Code	Spinal Nerve
C_1	First cervical nerve of spinal cord
C_2	Second cervical nerve of spinal cord
C_3	Third cervical nerve of the spinal cord
C_4	Fourth cervical nerve of the spinal cord
C_5	Fifth cervical nerve of the spinal cord
C_6	Sixth cervical nerve of the spinal cord
C_7	Seventh cervical nerve of the spinal cord
C_8	Eighth cervical nerve of the spinal cord
T_1	First thoracic nerve of the spinal cord
T_2	Second thoracic nerve of the spinal cord

APPENDIX B

Brief Description of Instrumentation for Measuring and Recording Force-Time Histories for Hand Grasping and Flexion and Extension in the Wrist and Elbow

The device is a bipod-mounted pistol grip (4-1/2 x 2 x 1 inch) spring loaded with three internally mounted piezoelectric crystals. It is expected that this will be a measure of grasping ability for sudden short grasps and/or prolonged holding. The lead zirconatetitinate crystals used in the prototype model have been dead-weight tested to 500 pounds with no appreciable non-linearity in electrical response.

Output from the pressure transducers is fed into a battery operated charge amplifier which is also used to drive a 2-channel pen recorder. The amplifier and recorder are housed in a portable carrying case weighing approximately 10 pounds and with an overall dimension of 12" x 8" x 6".

For flexion data, removable swivel screws are attached to each end of the hand grip. The swivel screws rotate in a U-clamp which, in turn, is affixed to the table or a rigid surface. The felt-backed nylon cords which are affixed to the pistol grip are grasped in the palm of the hand or slipped over the wrist. Pulling the cord orients the measuring device parallel to the direction of the flexion force. Removing the cord and pushing on the face of the grip will produce extension measurements.

APPENDIX C

Ballistic Missile Information

Weapon

Caliber: ____________

Type: ____________

Brand Name: ____________

Disposition of Weapon: ____________

Missile

Recovered: Yes ______ No ______

Brand Name: ____________

Disposition of Missile: ____________

____

Weight: (grams or grains) ____________

Dimensions: ____________

Situation

Distance between weapon and victim: ____________

APPENDIX D

Behavioral Information

Activity at moment of wounding:
(check)

Running ________

Standing ________

Sitting ________

Lying ________

Other (explain) ________ ____________________________

Remained conscious:

Yes _____ No _____

If no, was unconsciousness,

Immediate? _____

Later? _____

If later, about how long __________ mins.

Conscious responses:

A. Psychological:

Highly excited? ______

Stayed calm? ______

B. Physical:

		Remarks
Started fighting	______	____________________
Ran	______	____________________
Walked	______	____________________
Stood up	______	____________________
Fell to ground	______	____________________

APPENDIX E

Clothing and External Wound Tract Information

Description of Clothing Damage on Victim:

	Material	Description of Damage (Hole size, etc.)
Overcoat:		
Jacket:		
Shirt:		
Undershirt:		
Pants:		
Shorts:		
Other:		

Wound Extrance:

	Location(s) of Penetration(s)	Dimensions of Penetrations
(a)		
(b)		
(c)		

Wound Exit:

	Location(s) of Exit Hole(s)	Dimensions of Exit Hole(s)
(a)		
(b)		
(c)		

Clinical Subsystems

Name or Code No. ______________________ Age: ____ Race: __________

Hospital No. ______________________ Sex: _____

Date of Admission: __________

Date of Discharge: __________

Accident Room:

Time of Shooting: ______________________

Time of Arrival: ______________________

TIME

Clinical Observations	Acute Phase: Every hour	Control Phase: Every 4 hours	Recovery Phase: Once a day
Temperature:			
1. Cardiovascular Subsystem:			
Blood Pressure			
Pulse Rate			
Heart Sounds			
Electrocardiogram			
Laboratory Measures:			
Hemoglobin			
Hematocrit			
Red Blood Count			
White Blood Count			
Eosinophil Count			
Coagulation Time			
Sodium			
Potassium			
Chloride			
Proteins (Total)			
Estimated Blood Losses			
Volume Replacement			
Special Studies			

Clinical Observations (Cont'd)

2. Respiratory Subsystem:
    - Rate of Respiration
    - Breath Sounds
        - Right lung
        - Left lung
    - Palpation
        - Right lung
        - Left lung

    Laboratory Studies
    - X-Ray
    - Special Studies

3. Neurological Subsystem:
    - State of Consciousness
    - Mental Responses
    - Sensory Responses
    - Motor Reflexes
    - Cranial Nerve Responses

    Laboratory Studies
    - X-Ray
    - Electroencephalogram
    - Spinal Fluid

    - Special Studies

4. Skeletal Subsystem:
    - Fracture(s)
    - Dislocation(s)

    Laboratory Studies
    - X-Ray

    - Special Studies

<u>Clinical Observations</u> (Cont'd)

5. Muscular Subsystem:
   - Penetration/Perforation
   - Edema or Hematoma
   - Flacidity or Rigidity

   <u>Laboratory Studies</u>
   - X-Ray

   Special Studies

6. Gastrointestinal Subsystem:
   - Vomiting
   - Hematemesis
   - Condition of Bowel Sounds
   - Organ Palpation

   <u>Laboratory Studies</u>
   - X-Ray

   Special Studies

7. Genitourinary Subsystem:
   - Presence/Absence of Blood in Urine
   - Urine Volume
   - Urine Analysis

   Special Studies

8. Endocrine Subsystem:

   Special Studies

PROCEDURES:

THERAPIES:

APPENDIX G

Effector Measurements
(Upper Limbs Only)

	Admission	+4hrs.
Hand Grasping:		
Sudden maximal effort:	______	______
Prolonged grasp (loading to be specified):	______	______
Grasping follower exercise (to be specified):	______	______
Wrist Flexion:		
Sudden maximal effort:	______	______
Prolonged flexion (loading to be specified):	______	______
Wrist Extension:		
Sudden maximal effort:	______	______
Prolonged flexion (loading to be specified):	______	______
Elbow Flexion:		
Sudden maximal effort:	______	______
Prolonged flexion (loading to be specified):	______	______
Elbow Extension:		
Sudden maximal effort:	______	______
Prolonged elbow extension (loading to be specified):	______	______

APPENDIX H

Information on Normals

The formats anticipated for normals are duplicates of those used in Appendix F and Appendix G, i.e., wherever clinical information is sampled there is an assumed need for normal values. These values will either be drawn from similar populations and/or from those victims who survive the injuries and are considered to be normal again.

APPENDIX I

Anthropometric Information

Age: ______ Sex: ______ Weight: ______ Race: ______

Physical Measurements:

Height: ______

Weight: ______

Head circumference (forehead level):	______
Vertical distance (top of head to bottom of mandible):	______
Neck length (tip of hyoid bone to suprasternal notch):	______
Neck circumference (at midpoint of neck length):	______
Chest circumference (at nipple line):	______
Abdomenal circumference (umbilical level):	______
Hip level circumference (level of iliac crest):	______
Sternal notch to symphasis publis:	______
Spinous Process of C-7 to coccyx:	______
Width of shoulders (acromion to acromion):	______

	Right	Left
Acromion to radial epicondyle:	____	____
Midarm circumference:	____	____
Radial epicondyle to radial styloid process:	____	____
Radial styloid process to tip of middle finger:	____	____
Midforearm circumference:	____	____
Wrist circumference:	____	____
Anterior iliac spine to upper patella:	____	____
Upper patella to sole:	____	____

APPENDIX I
(cont'd.)

	Right	Left
Midcalf circumference:	______	______
Circumference at patella:	______	______
Midleg circumference:	______	______
Ankle circumference:	______	______
Length of foot:	______	______

PROBLEMS IN THE DESIGN OF STATISTICS - GENERATING WAR GAMES

William H. Sutherland
Research Analysis Corporation, McLean, Virginia

A newspaper headline last week said "Helicopters Crash in War Games." Now since I am billed as presenting a problem to you under the title "Problems in the Design of Statistics Generating War Games," the headline makes me hasten to tell you that the war games I'm concerned with are not at all like the newspaper version, and that whatever statistics that kind of war games generates in the form of crashes, such statistics have little in common with the kind I wish to talk to you about.

The kinds of war games that are of concern are played for Army research purposes. They are two-sided, somewhat formal exercises-played indoors using maps and often using computers. They are of a size and complexity measureable in tens of man months of play and analysis effort (an expensive kind of effort as operation research studies go, but seldom, I suspect, having costs comparable with the kind of war games in the newspaper headline). In our games records are kept of the details of play, and from these records statistics are derived. The statistics of the title, then, concern not real-world helicopters or troops or guns, but do concern the helicopters or troops or guns which the gamers have in mind's eye as they play. Battle results are found by applying rules, not bullets, and sometimes the battle results are made to depend on random numbers.

The players usually have a good deal of freedom of action tactically, and it is in this sense that I take the liberty of considering war games to be experiments. (Certainly the games do have this in common with experiments: we never know how they are going to come out.)

However, they are unlike most experiments in that one would not ordinarily expect to be able to repeat the initial conditions with strict exactness. If one were to try, with say the same players, they would no longer be in the dark about their adversary, because of what they had learned in the first game. If one tried a second set of players, they would necessarily have different tactical experience inside their heads to begin the game with.

Now after telling you that the tool--war gaming--is one that in one sense defies replication, let me ask you about a problem that war gamers face, which it seems to me, can be discussed partly in terms of what replication would show if indeed replication were possible. The problem presented itself in the course of a Research Analysis Corporation study on the use of war gaming as a research tool. It is: when should the war game designer use random numbers in the game, and when should he avoid using them? As you will see, the guidelines we have are only qualitative and what appear to us to be common sense. It would help if we had better guidelines, and this is the problem which I ask the panel to consider. So much for the introduction. What I have to say is in three sections: (1) the reasons for using random numbers; (2) the appropriateness of using average values versus random numbers; and (3) the effects of random numbers on interpreting game results.

1. REASONS FOR USING RANDOM NUMBERS. Games often make use of random numbers as a way of deciding details of combat. The use of random selections from previously determined or estimated probabilities serves two main kinds of function: to represent the chance nature of warfare and to keep players from having an inappropriate knowledge of their opponents.

As for any model, the chance nature of warfare can of course be only imperfectly represented. Only a few of the most important chance events can be selected to be part of a game. I suppose that this statement will seem a little like Alice-in-Wonderland to this audience for many of the papers presented in this conference the problem seems to have been to nail down chance events which wander uninvited into the problem - here we are in effect dragging them in by the heels. Of these chance events the use of random selections can be considered as a means of at least roughly representing the consequences of groups of causal sub-events. These sub-events result in the "main" event which is being represented. The sub-events--the direct causes--may be impractical or impossible to know, or they may in the game be unnecessary to know. As an analogy, the causes for the small deflections of a bullet fired from a gun in a test stand may well be impracticable to know; they depend on such matters as changes in air density along the patch of successive bullets or slight and unknown asymmetries in the loading of the powder. For many practical purposes, it is not necessary to know the causes. So also in

war games a spread of outcomes for an event can be used without specifying the causes. Thus random numbers can be used in games to represent events whose causes are not thoroughly understood or sometimes to simplify considerations which would otherwise be in too great a detail considering the scale of the game. For example, in a recent game the outcome--success or failure--of a minor raid against a logistics installation behind enemy lines was represented by a simple draw. Detailed consideration of the complex of factors for and against the raiders was not appropriate considering the probably minor effects of the raid on the overall game outcome.

The second class of use for random choice from distributions is to keep the opposing players suitably in the dark as to the exact capabilities of their opponents, and thus add to the realism of the game. Without random factors a player might be able to work the model or formula backwards and find enemy strengths. So randomness makes the players' decisions more like what they would be in real life.

Parenthetically, it has been observed that such use of random numbers, by making the incidents of a game less predictable, makes play more interesting to the participants. This contributes to the intensity and involvement which seems to characterize games, and which, one hopes, may occasionally result in an otherwise routine tactic being replaced with a brilliant one which may alter the concept being studied.

2. APPROPRIATENESS OF AVERAGE VALUES VS RANDOM SELECTIONS. Suppose that in a particular theater rain, if it occurs, has a strong effect on operations, but that it occurs only say three percent of the time, and without any repetive pattern (i.e., in a way that is reasonably represented as random). A single game is being played, and the random number generator indicates that it is to rain on some important day of the game. Is it appropriate to play that day according to rainy day rules? To do so would make the play for that day "non-typical"; to disregard the weather could be criticized as being unrealistic. So one subquestion to ask about random numbers is--when is it appropriate, for matters like weather effects to use "typical" values (i.e., dry weather); when to use average values (i.e., 97% dry); and when should the variations be randomly selected?

No hard and fast rules come to mind immediately, but certain observations on the subject can be made. The first is elementary. It is that there is little point in using the variations--which of course complicate the game--unless the factor itself is important to the game. If in the example rain had small significance, average values for its effects would certainly be sufficient.

Beyond being important in its general effect on the game output, though, further aspects of its importance need to be examined. One is whether the variation itself has an impact. If the effect being considered does not tend to do what I call "weight" the game results, then there may be no point in using any but an average value even though the overall effect is large.

Let me explain. One concept of the reason for playing war games, as opposed to say OR analytic studies is that the play permits examination of certain interactions that none of the other methods seem able to examine. Specific aspects of the interactions between

weapons,
tactics, and
environment,

can be studies in a combined manner in a game.

Let us take an example. As part of a recent war game at Research Analysis Corporation two competing antiaircraft systems were compared in the role of defending against armed helicopters. The particular tactics used by the helicopters included hiding behind available ridges and then coming in fast. The particular hilly terrain of the game limited the range at which the helicopters could be acquired as targets. For these tactics, then, and this terrain the more sophisticated and expensive weapon was not usable at the long ranges at which it was effective. The less complex weapon did nearly as well. The statistic used for the comparison was simply the number of helicopters shot down per helicopter sortie. In general, then, the game may be looked at as a method of examining the interactions between the weapons and the terrain, and the weapons and the tactics, or for that matter, any of the three as affected by the other two. Quantitative weights are implicit

in the play of the game and can be thought of as contributing to the relative use and relative effectiveness of the weapons or tactics under the game conditions. The game provides information and insights concerning such interactions, on the basis of such implicit weighting. The combining of tactics, weapons capabilities, and surrounding, may thus be thought of as being a product or output beyond the previous knowledge of the individual inputs. These are the experimental results with which we work. The game, in my viewpoint, is a particularly suitable and unique tool for Army use, partly because it can study such interactions.

To return to random numbers, there is, I suggest, little to be gained in using them to produce variability in some computational input or output unless such a range of numbers is tied to this weighting. The nature of the tie-in is, as far as I know, an unexplored area and one to which you may be able to suggest approaches.

3. INTERPRETATION OF GAME RESULTS - NUMBER OF RANDOM NUMBERS. But in order to do so with better insight, we might also look at the effects of using random numbers on the interpretation of game results. As I indicated, any single play would have to begin a little differently from any other play, and would then continue differently, even though the general circumstances of the game were similar. This comes about because the decisions of (all too) human players are involved, but also because of random number use. While the overall variability in results, which makes interpretation difficult, cannot really be separated into the two causes, we can for the time being assume that the players could be given only limited choices, and we can concentrate on the variability caused by random numbers. In effect this is what does happen in some computer simulations. Consider the number of times a random number is chosen for a particular purpose in the course of a game. In practice this varies greatly from one game to another. Some use random selections literally thousands of times in one play: others have few, and one two-sided exercise that was called a game did not use any. Let us examine a little the consequences of the use of different numbers of random selections. If the purpose, for example, was to find the battle outcome for a single highly important battle, and the random choice was made just once, then two quite different possible results could happen. In statistical terms, the spread of potential results would be large. (Incidentally what would be learned from such a game would be little.) On the other hand, if a long series of battles were fought, each of which

was equally important, and in which the probability of the different possible results did not change between battles, then statistically the relative spread of possible overall results would be small. Of course war games really do not present a picture of a large number of exactly similar events decided on a probability basis in the way described. Still, they presumably behave somewhat as if they did, and thus to a limited extent the spread of results can be discussed as a statistical matter. The limitations include this: that the measure of outcome is indeed a statistic--"winning" or "losing" is probably not such a measure. Secondly, for some mathematical considerations the measure must be an average. Certainly not all the statistics with which we are concerned are averages. But, to the degree that the measure of outcome used is a statistical average, the greater the number of random events that are applied directly to this outcome, the smaller the standard deviation of the result. The hypothetical universe of means we are speaking of gets relatively narrow as more choices are put in. Its standard deviation would vary inversely with the square root of the number of random choices made. Thus certain aspects of those games which repeatedly make a very large number of draws may be thought of as giving the same results as if average values had been used in the computations.

A complication involved in warlike simulations is that the entities one deals with (units or weapons) may be destroyed or eliminated. Thus one has a decreasing set of pieces to work with, and the relationship between the opposing pieces can well be changing as the work progresses. Thus, if a statistic is used like "the average number of Red entities destroyed per Blue counter-entity," such an average could be quite different at the beginning of a game than toward the end, when perhaps each side would at least need to do more hunting to find the enemy.

What should a game designer do then? Should he try to cut down the quantity of random numbers his game is to use, or try to increase them in the hope of reducing the deviation of possible results? No general rule seems evident at this time.

THE FUTURE OF PROCESSES OF DATA ANALYSIS*

John W. Tukey
Princeton University and Bell Telephone Laboratories

I am here to speak of the near-middle future of the processes of data analysis. This can only be done by indirection, since any processes that serve as examples must be drawn from the present of near future, but we can use such examples to illustrate what may be hoped to be broadly-applicable principles of continuing importance.

In general, the future of processes of data analysis is rosy, but it is not yet clear how fast the sun is rising. The modern computer has offered us many opportunities -- far more than we have seized -- and there have been many more opportunities for innovation that do not require a computer than we have seized. Looking at the last decade or two, it is clear that we have made much progress -- but we cannot be content with the rate at which we have gone. Will we do enough better in the future? Will we try to find approximate (or even crude) answers to more pressing problems, or exact answers to problems of limited (or non-existent) relevance? Who can say? (For a more historical and less specific discussion, see Tukey 1965.)

1. SOME PRINCIPLES. To point toward the near-middle future, we begin by stating a number of broad principles concerning the processes of data analysis (a phrase that ought to be construed as including the thoughts of the analyst of data as well as his manipulations) which we expect to retain their importance:

Two major aspects of such processes will continue their great importance:

(1) the essential erector-set character of data-analysis techniques, where any 2, 3 or 4 techniques are likely to be combined without warning,

(2) the steadily decreasing cost (and a so-far only slowly increasing ease) of computation, which is reflected in an ever-increasing

---

*Prepared in part in connection with research at Princeton University sponsored by the Army Research Office (Durham).

emphasis on computer usage and an ever-growing role of computer-unique contributions and processes.

Disproportionately rapid expansion will continue to repair past deficiencies in:

(3) graphicality and informality of processes of analysis,

(4) graphicality and incisiveness

(5) flexibility and fluidity

(6) empirical discovery of techniques

(7) focusing and parsimony.

In support of these improvements, our conceptual frameworks will give more and more attention to:

(8) doing the approximately right, rather than the exactly wrong (including dropping tight specifications as rapidly and generally as we may).

(9) using umbra-penumbra model pairs and other simultaneous (rather than alternative) model combinations.

(10) making the relation of estimator and target a two-way street.

And the day will yet dawn when:

(11) there will be one or more programming systems appropriate to data analysis.

As they stand, these principles are mainly unexplained words, requiring both examples and discussion to make then more understandable.

2. A GROUP OF EXAMPLES. Let us turn, then, to a group of examples, to instances of specific areas where progress is current. (I am sure my selection is biased, but this is only to be expected.) These

examples have not been selected to match the principles in a one-to-one way, instead each has been chosen to illustrate a few principles, with an attempt to illustrate each principle more than once. (Unfortunately, principle 11 cannot be adequately illustrated.)

Table 1 lists the examples and indicates their closest relations to specific principles.

I

THE EXAMPLES

3. NEWER APPROACHES TO TYPICAL VALUES. It has long been recognized that samples from distributions whose tails straggle more than those of a Gaussian were not well summarized by an equally weighted arithmetic mean. The procedures suggested by Jeffreys (1938, see also Newcomb 1886) for large samples have been occasionally implemented (e.g. Hulme and Symms 1939). More recently (i.e., Tukey 1960, Hodges and Lehmann 1962, Tukey and McLaughlin 1963) it has been recognized that what is needed is not so much a large-sample technique carefully bent to fit the particular distribution at hand, but rather techniques which provide relatively high efficiency over a wide range of distributions -- techniques that are (approximately) robustly efficient as well as being (approximately) robustly valid.

Much is being done in this area, and we shall soon have not only a body of asymptotic theory (Lehmann 1963a, b, c, 1964, Huber 1964, Buckel 1964) but an array of directly useful techniques (Hodges and Lehmann 1962, 1963, Høyland 1964, Dixon and Tukey 196?).

The problem of typical values in the plane, and in higher dimensions, is not so simple, since there is no obvious affine-invariant generalization of the notions of order statistics, which have played a central role in the one-dimensional case. Gentleman (196?) is tackling this problem from the point of view of minimizing p-th power deviations. Elashoff and Bickel (1964") are investigating Winsorizing and trimming. Soon we may expect working tools for this case, too.

Extensions to many other problems are obviously needed, and can be expected to occupy both asymptotic theorists and practical-technique designers over a considerable period.

TABLE 1

Title of example	Principles to which closely related: 1	2	3	4	5	6	7	8	9	10	(Section in which discussed)
Newer approaches to typical values	x					x			x		(3)
New dissections of factorial tables	x						x	x			(4)
Spectrum-like techniques		x	x	x		x				x	(5)
Unrestrained monotone transformations	x	x		x	x						(6)
Ordered plotting			x	x							(7)
Hanging rootograms			x								(8)
Deomnibusing							x	x			(9)
The jackknife								x		x	(10)
$\widehat{\text{Var}}$ for weighted means	x								x		(11)

4. <u>NEWER DISSECTIONS OF FACTORIAL TABLES</u>. For a long time only two dissections of data arranged in a two- (or more-) way table were common in data analysis. Both of these were almost always left implicit rather than made explicit. I refer, of course, to the additive decomposition, whose two-way form is

$$y_{ij} = y_{..} + (y_{i.} - y_{..}) + (y_{ij} - y_{..}) + (y_{ij} - y_{i.} - y_{.j} + y_{..})$$

that underlies the analysis of variance for crossed and nested factors, and the multiplicative decomposition, whose two-way form is

$$n_{ij} = n_{++} \left(\frac{n_{i+}}{n_{++}}\right) \left(\frac{n_{+j}}{n_{++}}\right) \left(\frac{n_{ij} n_{++}}{n_{i+} n_{+j}}\right)$$

that underlies the chi-square test for independence in contingency tables.

The cases where the labels of the columns, or the labels of the rows, or both, are at least ordered (and perhaps even relevantly quantitative) are important and deserve much attention. They are not, however, part of the subject we wish to discuss here. Our immediate concern is with decompositions other than the usual ones which can be carried out on <u>any</u> two-way (or more-way) table.

Among the earliest of these was the separation of one degree of freedom for non-additivity (Tukey 1949) in which the "row" and "column" parts of the usual decomposition were used to identify and separate part of the "interaction" parts. Further discussion (Tukey 1955, Scheffé 1959, Elston 1961), some generalizations (Ward and Dick 1952) and various modifications of [illegible] Harter and [illegible], [illegible]del 1959, Mandel 1961) followed.

The apparent needs of specific data analysis produced an extension along the lines of the "vacuum cleaner" (Tukey 1962) which does not function well in practice without the aid of some preliminary preparation (e.g. FUNOR-FUNOM, see Tukey 1962). This is only one of a branching family of alternatives that are still unexplored.

Some directions in which we ought to go are clear, but the details of tools and formulations are far from settled. We need to dissect a two-way table in more parts than the four indicated above. It will sometimes suffice to have:

(a1) An over-all contribution.
(a2) Column contributions.
(a3) Row contributions.
(a4) Unusual cell contributions.
(a5) Routine cell contributions.

As well as being important on their own, such dissections clearly have a close relation to the problems of Section 3.

Except for the smallest tables, it is likely to be necessary to go further, dissection row and column behavior into the unusual and the routine, just as for cell contributions. In either case, we will be prepared for both of these extremes:

(b) row and column effects clearly visible above a "noise" of routine cell contributions,

(c) a few cells deviating widely from all the others, which show no pattern of variation (including none by row or column).

We will be prepared for either extreme, since we shall be prepared for any mixture of these extremes.

We are here at a very early stage in the gaining of understanding. We have had some experience in the identification of unusualness, but we undoubtedly have much to learn. Once we are in reasonable shape for two-way tables, there are many ways to go.

5. SPECTRUM-LIKE TECHNIQUES. The application of Fourier methods to data gave rise to useful results in the simplest cases (e.g. Whittaker and Robinson 1924, Bartels 1940). The modern era in this area begins with the recognition (Bartlett 1950, Tukey 1950) that "white noise" is almost always a foolish null hypothesis, and that "white noise plus a few sharp lines" was an equally poor alternative hypothesis. Attention was first directed toward such questions as

consistent estimation of spectrum density (which the writer finds quite uninteresting, since he never saw even an approximately infinite amount of time series data) and variability (under Gaussian assumptions) of estimates of averaged spectrum densities. Later developments have emphasized the importance of keeping close touch with the average value of one's spectrum estimates and the advisability of introducing a variety of new techniques in order to approach the specific problems that are important in the specific application. (See Technometrics 1961 for a general introduction, including complex demodulation, see Akaike 1962 for misbehavior of the autocovariance function, see Akaike and Yamamouchi 1962 for practical problems in the use of cross-spectra, see Hasselman, Munk and MacDonald 1963 for the bispectrum, see Bogert, Healy and Tukey 1963 for the cepstrum, cross-cepstrum, pseudoautocovariance, and related concepts, etc., see MacDonald and Ward 1964 for interesting prediction-studying techniques.)

The two beliefs, both quite erroneous in the writer's view, that have contributed the most to delays and inadequancies in the use of spectrum analysis have been:

(a) A belief that, in using spectra, one ought to be concerned only with Gaussian situations.

(b) A belief that, in using spectra, one ought to be concerned only with stationary situations.

It is true that average value and spectrum only complete the specification of an ensemble of time series if we know more, say that the ensemble is Gaussian. This is, however, no more than the analog of the (equally correct) statement that average value and variance only complete the specification of a distribution when we know more, say that the distribution is Gaussian. (We do not, however, confine our use of variance to Gaussian distributions.)

In the absence of a suitably mathematical formulation and treatment of spectra for nonstationary ensembles, there has been an unfortunate tendency for some workers to feel that spectrum techniques should only be applied to situations of apparent stationarity. In practice, this can be quite foolish, as Munk and Snodgrass's discovery (1957) through their nonstationarity, of the weak long-period ocean waves arriving on our

Pacific Coast from the Indian Ocean and beyond, illustrates. In theory, it is at best dubious, since if our universe should repeat itself every 10^{11} to 10^{12} years, the whole universe (with all its time series) may perhaps be thought of as stationary -- and who can deny such a possibility?

It may prove fortunate that a mathematical formulation of the non-stationary case is now at hand (Priestley 1965) which tells us to do for slowly changing spectra just what we have done for plausibly constant spectra.

In addition to the new types of quantities being introduced and used, we are in the middle of a change in the actual computing techniques used to process the data. Where once some subsequence of:

(c1) taper
(c2) prewhiten
(c3) form mean lagged products
(c4) apply lag window
(c5) Fourier transform
(c6) hann or hamm, etc.

was relatively standard, alternative approaches, involving more computations linear in the observations before the formation of squares of products involving the data, are in use or contemplation. Techniques using complex demodulation appear to involve very real advantages, and are already in routine use (M. D. Godfrey 1964*). Now that complete Fourier transformation for N observations requires only a few times $N \cdot \log_2 N$ multiplications rather than N^2 (Cooley and Tukey 1965), we may well see computational techniques develop which start by complete Fourier transformation of the entire data. (The spectrum-analytic character of these techniques will be revealed by what happens next to the Fourier coefficients and how the ultimate quantities are interpreted.)

In economics, spectrum analysis is currently being applied to the problem of seasonal adjustment, and as a consequence economists are again thinking about the difficult question of what seasonal adjustment is really supposed to do.

So far as one can now see, spectrum-like analysis is going to continue to ramify and develop at a substantial rate.

6. UNRESTRAINED MONOTONE TRANSFORMATION. The fight between those who feared the loss of knowledge that comes from analyzing unwisely expressed data and those who feared serious biasing of levels of significance and confidence would come from expressing the data in the way in which it seemed to like to be expressed is an old one, but one that has never reached the front pages. Partly this has been because changes in modes of expression have seemed unimportant. Partly, I fear, it has been because those who realized that, in practice, 100% and 200% improvements in efficiency come more frequently from such changes than from almost anything else the analyst of data can do once the data is taken, have not advertised this fact sufficiently.

Those who have sought better modes of expression have traditionally chosen some simple family of transformations, often $z = (y+c)^P$, and have tried to choose the few parametric constants wisely in each particular instance. (For a clear exposition of a highly developed form of this approach, see Box and Cox 1964.) As the techniques have become more explicit, the hope of their wider application has increased steadily.

All this continues to be important, but the pressure of a real need for better multidimensional scaling has brought about a computer-aided revolution. The work of Shepard (1962, 1963) and Kruskal (1964a, b) has shown how much can often be gained by letting the computer choose whatever monotone transformation of the original value will lead to the simplest analysis. The impact upon multidimensional scaling and factor analysis is already substantial. Kruskal's reanalyses (196?) of Box and Cox examples show that even a 3x3x3 experiment may be big enough for such an analysis to be fruitful. We can hope for similar progress in many other areas (although semi-classical results on "Maximalkorrelation" show that we cannot do it everywhere).

7. ORDERED PLOTS. The classical example of plotting observed values rearranged in increasing order is the use of "probability paper" to show the apparent Gaussianity, or absence thereof, of a sample of observations. This example is classical, but it is still surprising how many statisticians have had little contact with the technique.

The arrival of the half normal plot (Daniel 1959) introduced a major change into the analysis of unreplicated and fractionated 2^P experiments. The idea that a set of contrasts could be used to show forth the unusual

size of its largest values, if any of their sizes were truly unusual, is not a difficult one. It is perhaps surprising that it took so long to appear.

Later, the more general technique of "gamma plotting", in which two parameters require estimation rather than one, was developed and applied in a variety of directions (Wilk and Gnanadesikan 1961, 1964a, Wilk, Gnanadesikan and Huyett 1962).

Today, the problem of adapting these techniques to the general analysis-of-variance situation, where different mean-squares have different numbers of degrees of freedom is being actively attacked with interesting results (Wilk and Gnanadesikan 1964b, Wilk, Gnanadesikan and Lauh 1964).

As a consequence of this, the writer is convinced that we shall see a partial return of the pendulum, which has now swung from analyses guided only by the natural order of lines (and the relations between average mean squares) of the analysis of variance to analyses guided only by the relative sizes of the mean squares. I, for one, believe that we ought to expect attention to both considerations in well thought-through analyses, though in ratios differing widely from instance to instance. (Given a complete 2^{12}, for instance, whose 4095 contrasts behave exactly like a Gaussian sample, I would regard the fact that the 12 largest contrasts were the 12 main effects as nonaccidential and highly significant.)

Here, too, we can, I believe, lift the curtain of the future a little When I try, I see signs of plots of gaps (= spacings) among the ordered observations appearing alongside -- and even in partial replacement of-- the more classical plots of the raw ordered observations. Time will tell.

8. HANGING ROOTOGRAMS. This example is included to show that even among the simplest of graphical techniques there can be new and useful techniques.

The histogram, with its columns of area proportional to number, like the bar graph, is one of the most classical of statistical graphs. Its combination with a fitted bell-shaped curve has been common since the days when the Gaussian curve entered statistics. Yet as a graphical technique it really performs quite poorly. Who is there among us who

can look at a histogram-fitted Gaussian combination and tell, reliably, whether the fit is excellent, neutral, or poor? Who can tell when the fit is poor, of what the poorness consists? Yet these are just the sort of questions that a good graphical technique should answer at least approximately.

How can we do better? If we have observed n_i cases in the ith class, we know that the variance of n_i is reasonably proportional to its average values (at least so long as n_i is not a large fraction of the total number of cases, n_+).

If we are to do a reasonable job of assessing fit, we deserve to have roughly constant variance. We can do this by replacing n_i by $\sqrt{n_i}$, as we are well aware in other contexts. We can do the same here, at least for the case of classes of equal width. We have only to take the square root (of the height) of the fitted curve at the same time that we take the square roots of the counts.

Because of the simple identity:

$$\sqrt{\text{one Gaussian density}} = (\text{constant})\,(\text{another Gaussian density})$$

the picture will look much the same -- in the large -- a family of rectangles compared with a Gaussian curve, but now variability is nearly constant (at the price of giving up the principle of "equal area for equal count" which has real uses in other directions but few if any in connection with goodness of fit).

But we are still comparing the ends of a row of rectangles with a curve, something the human-eye-and-brain combination is less than perfect at. How do we improve matters here? We have only to say, carefully and precisely, what we have always done, in order to learn what we might better do. Classically, we have taken a stack of rectangles, fixed one end of each on a horizontal line and compared the other ends with a curve. It is not a great step to say: "Let us take our stack of rectangles, fix one end of each on a curve and compare the other ends with the straight line. Why did we not do it long ago?"

While we are about it, we might as well turn the picture over, letting the curve hang down, supporting the rectangles. This third change completes our path to the "suspended rootogram" in which the eye can do so much more for us. (Some viewers prefer to stop at the "hanging rootogram" stage.) Figures 1, 2, 3, 4 [at the end of this article] show successive stages in the progress from conventional histogram to hanging rootogram.

Theré are other simple things to do in the graphical area, as we shall learn as we take care to realize that graphs can and should, among other things, be used for diagnosis as well as naive exhibition.

9. DEOMNIBUSING. The first step in data analysis is often an omnibus step. We dare not expect otherwise, but we equally dare not forget that this step, and that step, and other step, are all omnibus steps and that we owe the users of such techniques a deep and important obligation to develop ways, often varied and competitive, of replacing omnibus procedures by ones that are more sharply focused.

The replacement of group comparisons by multiple comparisons has been one of the outstanding phenomena of the last decade and a half. It has raised many deep issues on which we are far from being completely agreed -- whose discussion would take more space than we can here provide. So we note here only that a full account of the short-cut methods using ranges both in numerator and denominators is at last appearing (Kurtz, Link, Tukey and Wallace 1965, 196?).

We note also that progress has also been made on the deomnibusing of contingency table chi-square.

The detection of differences in the effects of ordered treatments -- under circumstances where the effects, if any, may be expected to be directly -- or antithetically -- ordered has at last engaged the attention of technique manufacturers. Two competing approaches exist, about which all protagonists will agree that either one is to be preferred to the unwise use of a flabby group comparison. One procedure is developed in a framework of successive testing (Bartholomew 1959, 1961a, b); the other in a framework of single contrasts of maximizing the least sensitivity (Abelson and Tukey 1959, 1963). (The writer notes a continuing preference for the latter, based on what he regards as good reason. Again space bars further discussion.)

Still more recently, there is progress in the deomnibusing of "goodness of fit" tests, which have always had so omnibus a character. For small samples, or compulsory heavy grouping, we need not merely sum the squares of standardized (or, better, Studentized) deviations to find a chi square. As has long been known (e. g. Cochran 1954) we can introduce any convenient set of orthogonal comparisons, and evaluate the results as separately or jointly as we wish. In doing this, it should be our hope to concentrate the effects of fitting the curve to which the data is compared as thoroughly -- and into as few comparisons -- as reasonably may be.

In larger samples, particularly in the absence of grouping, one can go a long way toward the separation of "badness of fit" into three parts:

(a1) underestimated badness of fit, where the almost inevitable fitting of parameters has concealed any true badness,

(a2) systematic badness of fit, where the deviations are both interpretable and indicative of inadequacy of shape of model,

(a3) irregular badness of fit, often an indication only of inadequacy of simple random sampling -- no evidence of inadequacy of distributional shape.

Once this is accomplished, the introduction of the ideas underlying ordered plotting allows us to break new ground, to -- reasonably and sensibly -- inquire as to goodness of fit for many kinds of nonrandom samples <u>without</u> preassumption of what kind of nonrandomness is involved. Early trials of such techniques have had quite illuminating results (Quandt 1964, 196?).

These are only the beginning. Deomnibusing of all our usual omnibus procedures will do much to occupy both technique-manufacturers and philosophy understanders in the years just ahead.

10. <u>THE JACKKNIFE</u>. The "jackknife" procedure allows almost any of us to set approximate confidence limits on almost any results calculated from data which go a reasonable way toward revealing the variation whose likely effects are to be spanned by the confidence interval.

In its simplest form, the jackknife procedure assumes that

(a1) we have data, and a fixed procedure for extracting an interesting number (or numbers) from the data,

(a2) this procedure can be applied to varying amounts of data,

(a3) the data can be divided into r "pieces" of roughly equal "size",

(a4) this can be done in such a way as to make the differences from piece to piece "adequately reflect" the sorts of variation whose effects are to be spanned by the confidence interval,

(a5) the prototype case of "adequately reflect" is the sampling of r "pieces" from a very large collection of pieces, whose combined processing would, by definition give the right answer,

(a6) the results of the processing are not narrowly estimated, in the sense that no one piece has (and no very few pieces have) a dominating effect upon the result.

Given all this, to some reasonable approximation and according to some reasonable belief (which is all that one can ever truly demand) the analyst treats his data as follows:

(b1) Let y_{all} be the result of processing all the pieces of data together.

(b2) Let $y_{(i)}$, read "y-not-i", be the result of processing all but the ith piece of data (hence processing r-1 out of r of the pieces together).

(b3) Let y_{*i}, read "y-pseudo-i", be given by

$$y_{*i} = ry_{all} - (r-1)y_{(i)}$$

(b4) Let $y_* \pm t_{r-1} \cdot s_*$ be the mean of the y_{*i} and the confidence interval generated by a naive application of Student's to the y_{*i} (as if they were a sample).

The procedure is simple, the approximation is usually satisfactory, and the technique is applicable in very diverse and complex circumstances.

Happily this technique has begun to receive attention from some of those fitted to pinpoint some of its weaknesses and difficulties. In particular, there has been inquiry into the asymptotic behavior of the technique, especially where condition (a6) fails (Miller 1964). It is to be hoped that there will be more such studies -- and that their results will be correctly evaluated from the point of view of practice.

The cases where (a6) is most likely to fail are those in which a single order statistic, a median, a maximum, a minimum, or a few order statistics play an unusually important part. In some of these, particularly where medians and other inner order statistics are concerned, we have other means of assessing the stability of our answers that are adequately robust. In these cases we should clearly use these alternate procedures.

In others, often those involving maxima, minima, and ranges, it is clear that a properly assessed uncertainty for the quantity of interest will inevitably depend on such matters as the actual shape of the underlying distribution or distributions. Here robustness is impossible, and so is certainty of validity for any confidence procedure. It will often be true that the best that we can do is to use the jackknife in such situations, even though we know it may be fallible. It is usually better to have some idea of the uncertainty of our values rather than none. (No confidence interval will ever be computed from data in such a way as to include all possible sources of variation, since no body of data allows all possible sources to reveal themselves. A little more inadequacy will not be fatal.)

In cases where (a6) is not in question, the situation is rather similar. If there is available a robust special confidence procedure clearly applicable to the case at hand, by all means use it. Otherwise use the jackknife.

11. ESTIMATED VARIANCES FOR WEIGHTED MEANS.

(a) Given n uncorrelated observations y_i with the same average value and fixed finite variances

for which also

(b) the variances σ_i^2 of y_i are all equal, say to σ^2, it is well known that

$$s^2 = \frac{1}{n-1} \Sigma (y_i - \bar{y})^2$$

where

$$\bar{y} = \frac{1}{n} \Sigma y_i$$

is an unbiased estimate of σ^2 and that

$$s_1^2 = \frac{1}{n} s^2$$

is an unbiased estimate of the variance σ^2/n of $\bar{y}$.

It is known, but not to enough people, or clearly enough, that (b) has no part in the relation

$$(*) \qquad \text{ave } s_1^2 = \text{var } \bar{y}$$

which is a consequence of (a) alone.

We have here a simple example of an umbra-penumbra situation in which two models, one encompassing the other, are wisely considered simultaneously. The penumbra or outer model, here defined by (a) above, suffices for the validity of s_1^2 as an estimated variance $\bar{y}$ (in the sense that (*) then holds). The umbra or inner model, here defined by (a) and (b) together, ensures the optimality of s_1^2 as the unique quadratic function of the y_i that (i) satisfies (*) and (ii) minimises its own variance among the quadratics that do this.

The pattern here: "validity in the outer model, optimality in the inner" is but one of many possible patterns for simultaneous model pairs. It is however, one of the most important ones, one that needs much more explanation.

Suppose, for instance, that our concern is not with $\bar{y}$ but with

$$y_C = \frac{\Sigma c_i y_i}{\Sigma c_i}$$

where the $c_i \geq 0$ are fixed. Can we use the values of the c_i to determine a quadratic function $Q(y_1, y_2, \ldots, y_n)$ so that

(**) $$\text{ave } Q = \text{var } y_C$$

provided only that (a) holds and the c_i are as assumed? Certainly we can do this. We can, indeed, press right on, and find a Q which (i) satisfies (**) under (a) and (ii) minimizes its own variance under (a) and (b) combined. Nay more, we may replace (b) by

(c) the variances of the y_j are in known ratio, in that

$$\text{var } y_i = d_i \sigma^2$$

where the d_i are fixed and known.

For each choice of $\{d_i\}$, there will be a Q satisfying (**) under (a) and minimizing var Q under (a) and (c). This Q will, in fact, be different for different choices of $\{d_i\}$.

We could write down, in closed form, expressions for these Q's, but their detailed form is of far less concern to us than the facts that

(d1) we can have any of many umbras with a single penumbra

(d2) which umbra we choose can, sometimes, turn efficiencies topsy-turvy without affecting validity

(d3) the equally weighted mean seems to have no unusual roles; it appears to be just another weighted mean, the one, perhaps, for which certain formulas look simplest.

We need, and are inevitably going to get for ourselves, a very much wider collection of instances where the consequences of umbra-penumbra model pairs have been worked out, much to our illumination and advantage.

II

THE RELATION OF EXAMPLES TO PRINCIPLES

12. HOW THE EXAMPLES ILLUMINATE THE TWO MAJOR ASPECTS. The first (erector-set) principle is, according to Table 1, illustrated by:

(a1) newer approaches to typical values: where Winsorizing is combined with Student's t; where techniques developed for single samples are expected to be used directly or indirectly in simple and multiple regression and in all sorts of analyses of variance involving replication within cells,

(a2) new dissections of factorial tables: where we try to use both factorial and idosyncratic dissections at the same time; where we expect to build each new kind of dissection into more and more complex patterns,

(a3) unrestrained monotone transformation: which is rapidly propagating itself in cooperative combination with a wide variety of other techniques,

(a4) internally estimated variances for weighted means: where we learn how to do, knowingly, for weighted means what we have so long done, often unknowingly, for equally weighted means.

While not one of these is as striking as Cuthbert Daniel's unpublished injections of 2^{j-m} fractional factorial analysis into the calculations of multiple regression, or as striking as the technique-combinations that are, in practice, appropriate to a variety of complex bodies of data, they do offer solid illustrations.

The second principle (computation cheaper, more used, and more vital) is certainly well-exemplified above. Consider:

(b1) spectrum techniques: where hand-calculator work would be worth while for some of the most crucial instances, but where the cost

of hand computation, if it had to be paid, would keep us from many of the useful and illuminating studies we actually make.

(b2) unrestrained monotone transformation: where iterative computer algorithms lead us easily to ends not at all reasonably accessible by hand computation.

(b3) new dissections of factorial tables: where, when the still more effective techniques arrive, their feasibility will depend greatly, perhaps absolutely, upon the availability of computers.

(b4) ordered plotting: where much of the real push forward seems to be associated with the use of modern computer both to calculate and to make such plots.

(b5) deomnibusing: where, while the deomnibusing of goodness of fit is, in many instances, feasible without a modern computer, it is the availability of computer procedures that will make such techniques popular.

(b6) the jackknife: where one of the great advantages is that a well-written computer problem to do y_{all} can also be used to do y_{*j}, so that the cost of jackknifing is only a little more running time, but no extensive effort in programming or debugging.

While these are matters of technique manufacture rather than technique use, many of the new approaches to typical values are only accessible because of the modern computer (as when Monte Carlo techniques are required to find critical values, even when the underlying distribution is Gaussian).

There can be little doubt of the importance of the second principle.

13. HOW THE EXAMPLES ILLUMINATE THE FIVE AREAS OF RAPID EXPANSION. The first area where rapid expansion is trying to repair past deficiencies, principle 3, involves graphicality and informality, where the graph is used as an effective, but very informal, way of connecting the data to the human judgments that are going to be made about it, that constitute the reasons for its analysis. Here, Table 1 points out:

(a1) spectrum-like analysis: where a far larger share of judgments than many might suppose are in fact based on graphical presentations, informally examined.

(a2) ordered plotting: where a formal significance testing procedure, is largely replaced by an informal judgment made by those who look at the plot.

(a3) hanging rootograms: where we have striven to learn graphically and far less formally by seeking an approach to goodness of fit where a moderately wise man's eye will tell most of the story.

The second area, principle 4, involves graphicality and incisiveness and is quite distinct from the first, although the areas share graphicality and appear together in many techniques. At issue here is that grand property of many graphs: revelation of the unexpected through the simultaneous revealability of many possible deviations from neutrality. Table 1 directs our attention to:

(b1) spectrum-like techniques: where little peaks have often revealed new phenomena, as in Munk and Snodgrass 1957, or as in the detection and evaluation of the natural modes of vibration of the earth;

(b2) unrestrained monotone transformations: where the graph of the final monotone transformation is often quite revealing; where the structures resulting from multidimensional scaling often show unexpected properties.

(b3) ordered plotting: where we have learned that half-normal plots expose many kinds of interesting behavior other than the stray large values to detect which the plot was invented (Daniel 1959).

The third area, principle 5, involves flexibility and fluidity and deserves discussion in two rather separate pieces. Flexibility here refers to the existence of a wider variety of frameworks for analysis and inference, thus offering, on the average, a better match to the needs of the problem. Fluidity refers to the ability of single analytical procedures to respond in a very wide variety of ways to the apparent character of individual bodies of data. Clearly a continuous graduation from flexibility to fluidity is possible -- indeed many stages along this

gradation will be realized -- but for all that it still will pay us to try to make the distinction between choices made *in advance* by the analyst and choices made *en passant* by the data.

We have not done much to illustrate flexibility *directly* in our examples, yet most of them illustrate it indirectly, thus:

(c1) Our newer approaches to typical values are not yet focused into one form -- as they might some day be, where they completely fluidized. (As I trust they never will be, since I expect that the analyst will always be very often able to add information about the underlying distribution over and above that contained in a single small sample.) Shall we use trimmed means, Winsorized means, or Hodges-Lehmann median differences? If we trim or Winsorize, how far? We have not yet provided the user with the information most helpful in choosing answers to these questions, but we have begun to provide him the flexible kit of tools from which to choose.

(c2) There will be more than one choice among new dissections of factorial tables.

(c3) One can rightfully say that the modern phase in spectrum-like analysis comes from expanding our kit of tools beyond the serial correlation function and the periodogram (neither of which was really helpful).

(c4) The jackknife is a great aid to flexibility; in most situations it removes that grim complaint "But if we do *that* how can we compare the result with chance fluctuations?" and allows much freer choice of technique.

(c5) The ability to estimate variability for all weighted means with the same robustness as for the equally weighted case is a similar contributor to flexibility.

For the moment, the unrestrained monotone transformation is the outstanding example of complete fluidity. We face a challenge in finding others.

The fourth area of growth, principle (6), involves the empirical discovery of techniques, as opposed to their theoretico-mathematical

discovery. Here again we deal with a matter of amount rather than kind. It seems likely that no technique was developed solely empirically, without any "theoretical" insight at all, though many have been developed without any trace of a mathematical mode. At the other extreme, techniques based on rigid mathematical models, clearly-specified criteria, and vigorous optimization only gain credibility from some empirical support, whether of their hypotheses or of their functioning in practice.

Accordingly, our examples will tend to be ones that show a greater empirical content than most, ones whose developments are separated in amount, rather than kind, from those of most techniques. Table 1 cites:

(d1) newer approaches to typical values: where "trimming" and "Winsorization" came into being at least as much because of how they worked in practice as for any insight or theoretical argument: where the matching of denominators to numerators has come about by empirical comparisions based on tables of order statistic moments; where the critical values have often to be determined by Monte Carlo.

(d2) spectrum-like techniques: where one source of modern lag-windows was Hamming's observation that the points of an estimated spectrum for a single particular set of data would be improved by hanning; where the pseudoautocorrelation was suggested by a diffuse analogy with the cepstrum, and only the fact that it seemed to work made it plausible.

Though not quite a technique of data analysis, the near constancy of standardized 5% distances for Pearson curves (Pearson and Tukey 196?) is based upon Charles P. Winsor's wholly empirical discovery of the near constancy of the standardized 5% distance for chi-square.

The fifth growth area, principle 7, is one of focusing and parsimony. Some books on probability and statistics reveal that every sample (or other grouping of observations) is unusual in some way. (If only by how closely it matches a copy of itself.) It is rare, however, that the discussion carries on to the logical conclusions: First, that it is important to be restrictive in the kinds of unusualness to which one pays attention. Second, that one escapes this difficulty when one can focus all one's attention upon a single numerical aspect, or on a very few numerical aspects. Third, that once a fair number of such aspects are involved one is in a situation very like the unrestricted case and

that just how one divides his attention is of great importance. One can be wisely parsimonious with one of one's most valuable possessions, by focusing one's attention where this is most likely to be profitable.

Table 1 directs our attention to:

(e1) new dissections of factorial tables: where, instead of merely giving a single number to an inchoate mass of "interaction", we are striving to attend to very particular aspects, such as the single very unusual cell or indications that some other mode of expression will lead to a better approximation to additivity.

(e2) deomnibusing: in each of whose specific instances we are trying to improve our focusing, to learn about something identifiable and thereby to increase both the value of our knowledge and the chance of gaining it.

14. <u>HOW THE EXAMPLES ILLUMINATE THE AREAS OF INCREASED ATTENTION.</u> The first principle of increased attention, principle 8, calls for greater attention to being approximately right rather than exactly wrong. The hardest part of this, at least for the mathematician, is to admit that one is proceeding approximately -- even though it is hard to see how one can ever do better in the real world.

Table 1 directs our attention to:

(a1) new dissections of factorial tables: where we are seeking to ask the questions of greatest importance to us, even though their asking tends to destroy the neat, nice, manageable, null hypothesis which was the formal foundation for the classical asking of less useful questions; where our conclusion levels are going to become approximate; where there will be, for a period of years at least, no formal criterion to insulate us from the very real difficulties of picking a good technique.

(a2) deomnibusing: where we are again very willing to be approximate in the answering of more meaningful questions.

(a3) the jackknife: where by admitting that an approximate conclusion procedure can serve us, we have brought a very much wider range of techniques into the fold for which confidence, and significance, statements are at hand for use when appropriate.

The second principle of increased attention, principle 9, calls for more use of model-pairs and other "it might be A and it might be B and we must think about both together" approaches. The use of pairs of models as alternatives, as in the Neyman-Pearson account of hypothesis testing, is classical. (Pearson (1939) points out how much Student had to do with the recognition of its importance.) It is remarkable, by contrast, how little attention has been paid to pairs of models simultaneously considered. Perhaps this is because, in many instances, the use of simultaneous model pairs inevitably attracts attention to the deficiencies of a technique. In an optimality-validity umbra-penumbra situation, for example, emphasizing the validity of the technique in the penumbra cannot help reminding us that it is not optimum throughout.

Table 1 draws our attention to:

(b1) newer approaches to typical values, where Gaussian and crudely Gaussian underlying distributions provide umbra and penumbra that are used in varied ways: relatively good efficiency for the Gaussian and validity for all symmetric distributions; critical values set for the Gaussian (and approximately valid elsewhere) and moderately high efficiency (except for unseizable opportunities) anywhere near the Gaussian: etc.

(b2) internally estimated variance for weighted means: where the whole discussion is on an umbra-penumbra basis.

The third principle of increased attention, principle 10, calls for making the relation of estimator and estimand a two-way street. (See Tukey 1962, p. 10 and references cited there.) The mathematician wants the problem to come before the solution. But a good solution can often be recognized as such before we have identified one or more of the problems it solves. And a good solution may be good because it solves a problem other than the one as whose solution it is customarily derived.

Table 1 directs our attention to:

(c1) spectrum-like techniques: where much has been gained by asking what spectrum estimates actually do estimate, rather than by asking for asymptotic results which demand unreal amounts of data,

(c2) the jackknife: where the estimator is defined by a process, selected by what wisdom the analyst possesses, and the estimand follows after it, like the tail of a kite.

In each of these three areas of increased attention, if one goes through the uncited examples carefully, one will find each principle recurring again and again, though usually less explicitly. If one looks at the three areas in the right way, they seem to blur and move together into one.

If we look at all the principles, the same blurring appears, though not as obviously. There is a sense in which all these principles are "sisters under the skin".

III

THE CONCLUSION

15. SUMMARY. If we ask of the near future of processes of data analysis, one can predict three essentials:

(d1) greater realism,

(d2) greater effectiveness,

(d3) greater use of computers.

REFERENCES

(The final digit "?" indicates an unpublished paper, while an * identifies a personal communication.)

R. P. Abelson and J. W. Tukey 1959. Efficient conversion of nonmetric information into metric information. Proc. Social Statist. Section (Amer. Statist. Assoc.) 1959, 226-230.

R. P. Abelson and J. W. Tukey 1963. Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. Annals Math. Statist. 34, 1347-1369.

H. Akaike 1962. Undamped oscillation of the sample autocovariance function and the effect of prewhitening operation. Annals Inst. Statist. Math. (Tokyo) 13, 127-143.

H. Akaike and Y. Yamamouchi 1962. On the statistical estimation of frequency response function. Annals Inst. Statist. Math. 14, 23-56.

J. Bartels 1940 (2nd ed. 1951). Periodicities and harmonic analysis in geophysics. Chapter 16 (pp. 545-605, opening vol. 2) of Geomagnetism (by S. Chapman and J. Bartels) Oxford University Press.

D. J. Bartholomew 1959. A test of homogeneity for ordered alternatives I, II. Biometrika 46, 36-48 and 328-335.

D. J. Bartholomew 1961a. A test of homogeneity of means under restricted assumptions. J. Roy. Statist. Soc. Ser. B, 23 239-271 (discussion 271-281).

D. J. Bartholomew 1961b. Ordered tests in the analysis of variance. Biometrika 48, 325-330.

M. S. Bartlett 1950. Periodogram analysis and continuous spectra. Biometrika 37, 1-16.

P. J. Bickel 1964. On some robust estimates of location. (Abstract) Ann. Math. Statist. 35, 1403.

B. P. Bogert, M. J. R. Healy and J. W. Tukey 1963. The frequency analysis of time series for echoes: cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking. Chapter 15 (pp. 209-243) of (Proc. of the Symp. on) Time Series Analysis (ed. M. Rosenblatt). New York, John Wiley and Sons.

G. E. P. Box and D. R. Cox 1964. An analysis of transformations. J. Roy. Statist. Soc. Series B **26**, 211-243 (discussion 244-252).

W. G. Cochran 1954. Some methods for strengthening the common χ^2 tests. Biometrics **10**, 417-451.

J. W. Cooley and J. W. Tukey 1965. An algorithm for the machine calculation of complex fourier series. Mathematics of Computation **19** (in press).

C. Daniel 1959. Use of half-normal plots in interpreting factorial two level experiments. Technometrics **1**, 311-341.

W. J. Dixon and J. W. Tukey 196? Approximate behavior of the distribution of Winsorized t (Trimming/Winsorization 2). To be submitted for publication.

R. M. Elashoff and P. J. Bickel 1964*. Personal communication.

R. C. Elston 1961. On additivity in the analysis of variance. Biometrics **17**, 209-219. (Also abstract 688 at page 166.)

W. M. Gentleman 196? Robuts affine location estimation by minimizing pth power deviations. Ph.D. thesis (in progress) Princeton University.

M. D. Godfrey 1964*. Personal communication about programs in use at the Princeton University Computing Center.

H. L. Harter and M. D. Lum 1958. A note on Tukey's one degree of freedom for non-additivity. (Abstract 474). Biometrics **14**, 136-137.

K. Hasselman, W. Munk and G. MacDonald 1963. Bispectra of ocean waves. Chapter 8 (pp. 125-139) of (Proc. of the Symp. on) Time Series Analysis (ed. M. Rosenblatt). New York, John Wiley and Sons.

J. L. Hodges, Jr. and E. L. Lehmann 1962. Rank methods for combination of independent experiments in analysis of variance. Annals Math. Statist. 33, 482-497.

J. L. Hodges and E. L. Lehmann 1963. Estimates of location based on rank tests. Annals Math. Statist. 34, 598-611.

A. Høyland 1964. Evaluation of Hodges-Lehmann estimates (Abstract). Annals Math. Statist. 35, 938.

P. J. Huber 1964. Robust estimation of a location parameter. Annals Math. Statist. 35, 73-101.

H. R. Hulme and L. S. T. Symms 1939. The law of error and the combination of observations. Roy. Astron. Soc. Monthly Notices 99, 642-649.

H. Jeffreys 1938. The law of error and the combination of observations. Phil. Trans. Roy. Soc. Lond. A. 237, 231-271.

J. B. Kruskal 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, 1-27.

J. B. Kruskal 1964b. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29, 115-129.

J. B. Kruskal 196? Analysis of factorial experiments by estimating monotone transformation of the data. To appear in J. Roy. Statist. Soc. Ser. B.

T. E. Kurtz, R. F. Link, J. W. Tukey and D. L. Wallace 1965. Short-cut multiple comparisons for balanced single and double classifications, Part 1, Results. Technometrics 7 (in press).

T. E. Kurtz, R. F. Link, J. W. Tukey and D. L. Wallace 196? Short-cut multiple comparisons for balanced single and double classifications, Part 2, Derivations and approximations. Submitted to Biometrika.

E. L. Lehmann 1963a. Robust estimation in analysis of variance. Annals Math. Statist. 34, 957-966.

E. L. Lehmann 1963b. Asymptotically nonparametric inference: an alternative approach to linear models. Annals Math. Statist. 34, 1494-1506.

E. L. Lehmann 1963c. Nonparametric confidence intervals for a shift parameter. Annals Math. Statist. 34, 1507-1512.

E. L. Lehmann 1964. Asymptotically nonparametric inference in some linear models with one observation per cell. Annals Math. Statist. 35, 726-734.

N. J. MacDonald and F. Ward 1964. The prediction of geomagnetic disturbance indices. 1. The elimination of internally predictable variations. J. Geophys. Res. 68, 3351-3373.

J. Mandel 1959. The analysis of latin squares with a certain type of row-column interaction. Technometrics 1, 379-387.

J. Mandel 1961. Non-additivity in two-way analysis of variance. J. Amer. Statist. Assoc. 56, 878-888.

R. G. Miller, Jr. 1964. A trustworthy jackknife. Annals Math. Statist. 35, 1594-1605.

W. H. Munk and F. E. Snodgrass 1957. Measurements of southern smell at Guadalupe Island. Deep Sea Research 4, 272-286.

S. Newcomb 1886. A generalized theory of the combination of observations so as to obtain the best result. Amer. J. Math. 8, 343-366.

E. S. Pearson 1939. "Student" as statistician. Biometrika 30, 210-250 (especially pp. 242-243).

E. S. Pearson and J. W. Tukey 196? Approximate means and standard deviations based on distances between percentage points of frequency curves. (In preparation.)

M. B. Priestley 1965. Evolutionary spectra and nonstationary processes. J. Roy. Statist. Soc. Series B 27 (to appear).

R. E. Quandt 1964. Old and new methods of estimation and the Pareto distribution. Econometric Research Program Research Memorandum No. 10. Princeton University (submitted to the J. Amer. Stat. Assoc.).

R. E. Quandt 196? Statistical discrimination among alternative hypotheses and some economic regularities. J. Regional Science (in press).

H. Scheffé 1959. The Analysis of Variance (477 pp.) New York, John Wiley (especially pp. 129-134).

R. N. Shepard 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika 27, 125-140. II. Psychometrika 27, 219-246.

R. N. Shepard 1963. Analysis of proximities as a technique for the study of information processing in man. Human Factors, 5, 33-48.

Technometrics 1961. (Papers on spectrum analysis by G. M. Jenkins (133-166, 229-232), E. Parzen (167-190, 232-234), J. W. Tukey (191-219), T. H. Wonnacott (235-243) and N. R. Goodman, S. Katz, B. H. Kramer, M. T. Kuo (245-268)) Technometrics 3, 133-268.

J. W. Tukey 1949. One degree of freedom for non-additivity. Biometrics 5, 232-242.

J. W. Tukey 1950. The sampling theory of power spectrum estimates. (pp. 47-67) in Symposium on Applications of Autocorrelation Analysis to Physical Problems (NAVEXOS-P-735) Office of Naval Research.

J. W. Tukey 1955. Answer to query 113. Biometrics 11, 111-113.

J. W. Tukey 1960. A survey of sampling from contaminated distributions. Paper 39 (pp. 448-485) in Contributions to Probability and Statistics (edited by I. Olkin et al) Stanford University Press.

J. W. Tukey 1962. The future of data analysis. Annals Math. Statist. 33, 1-67.

J. W. Tukey 1965. The technical tools of statistics. The American Statistician (to appear, probably in April).

J. W. Tukey and D. H. McLaughlin 1963. Less vulnerable confidence and significance procedures for location based on a single sample (Trimming/Winsorization 3). Sankhya 25, 331-352.

G. C. Ward and J. D. Dick 1952. Non-additivity in randomized block designs and balanced incomplete block designs. New Zealand J. Science and Techn. B. 33, 430-435.

E. T. Whittaker and G. Robinson 1924 (4th, 1946). The Calculus of Observations (397 pp.) London, Blackie and Son.

M. B. Wilk and R. Gnanadesikan 1961. Graphical analysis of multiple response experimental data using ordered distances. Proc. Nat. Acad. Sci. USA 47, 1209-1212.

M. B. Wilk, R. Gnanadesikan and M. J. Huyett 1962. Probability plots for the gamma distribution. Technometrics 4, 1-20.

M. B. Wilk and R. Gnanadesikan 1964a. Graphical methods for internal comparisons in multiresponse experiments. Annals Math. Statist. 35, 613-631.

M. B. Wilk and R. Gnanadesikan 1964b. A probability plotting procedure for internal comparisons in a general analysis of variance. Unpublished invited talk presented at Royal Statist. Soc. meetings, Cardiff, Wales, in September 1964.

M. B. Wilk and R. Gnanadesikan and E. Lauh 1964. Scale parameter estimation from the order statistics of unequal gamma components. Unpublished memo., Contributed paper presented at IMS meetings at Amherst, Mass., in August 1964.

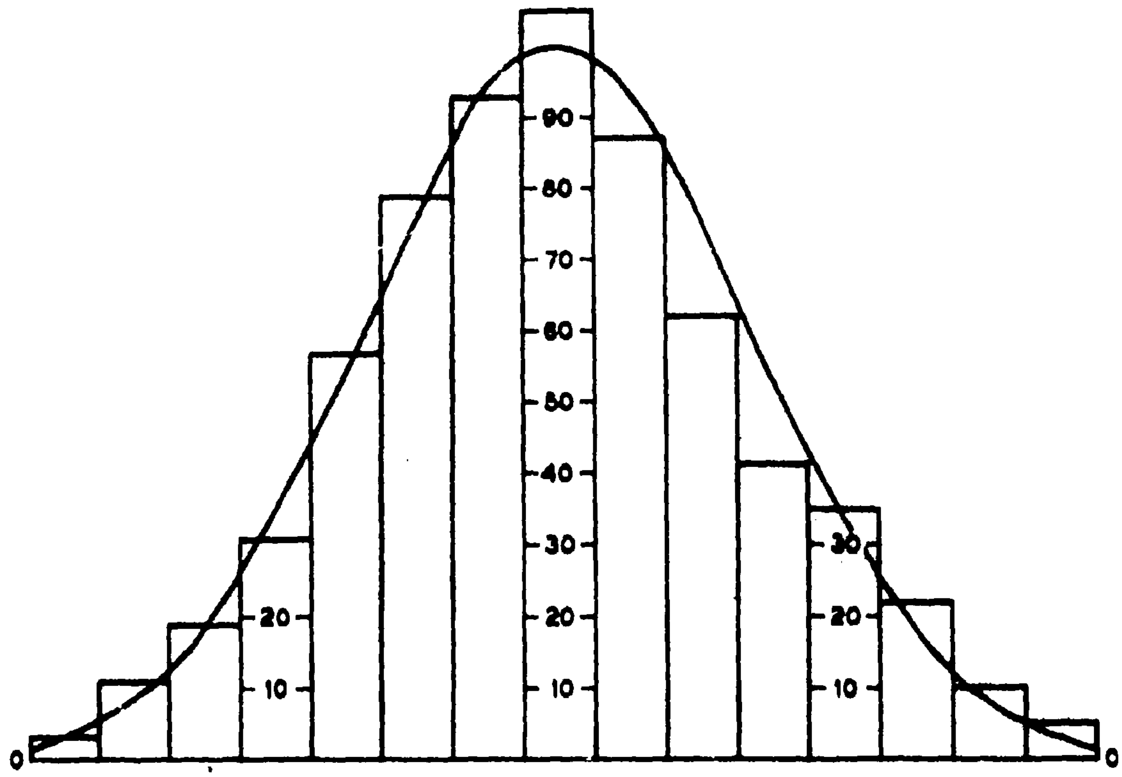


FIG. 1
HISTOGRAM

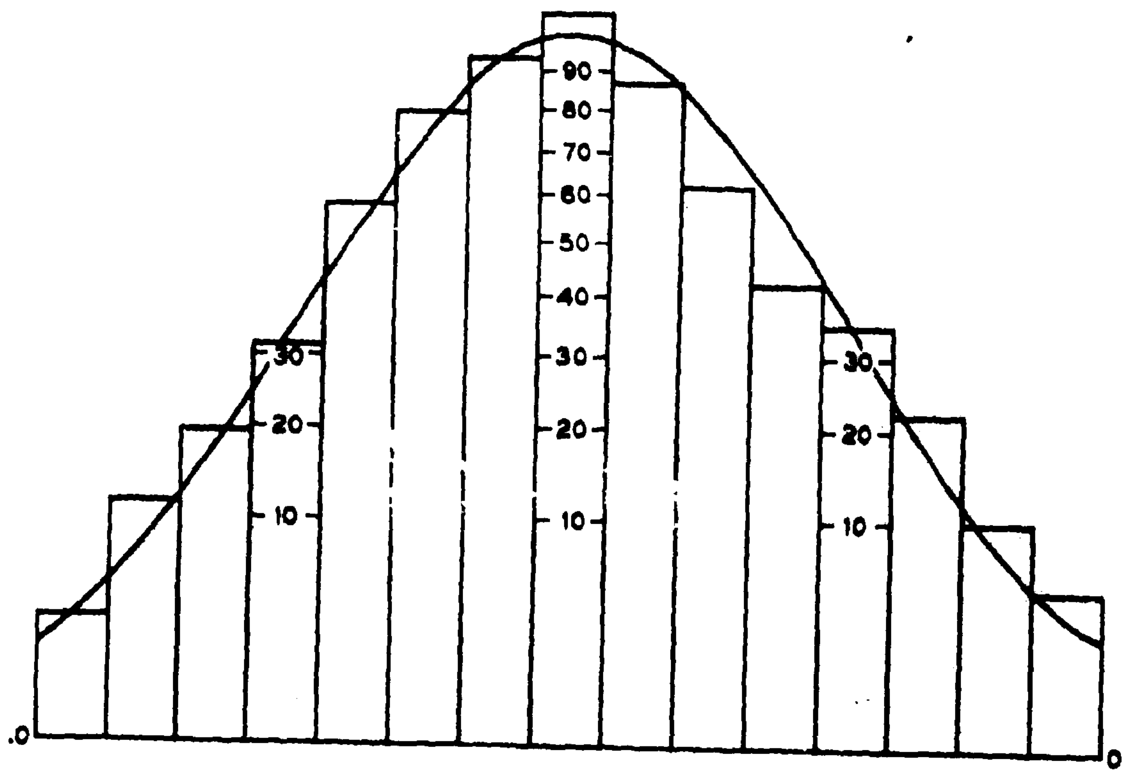


FIG. 2
ROOTOGRAM

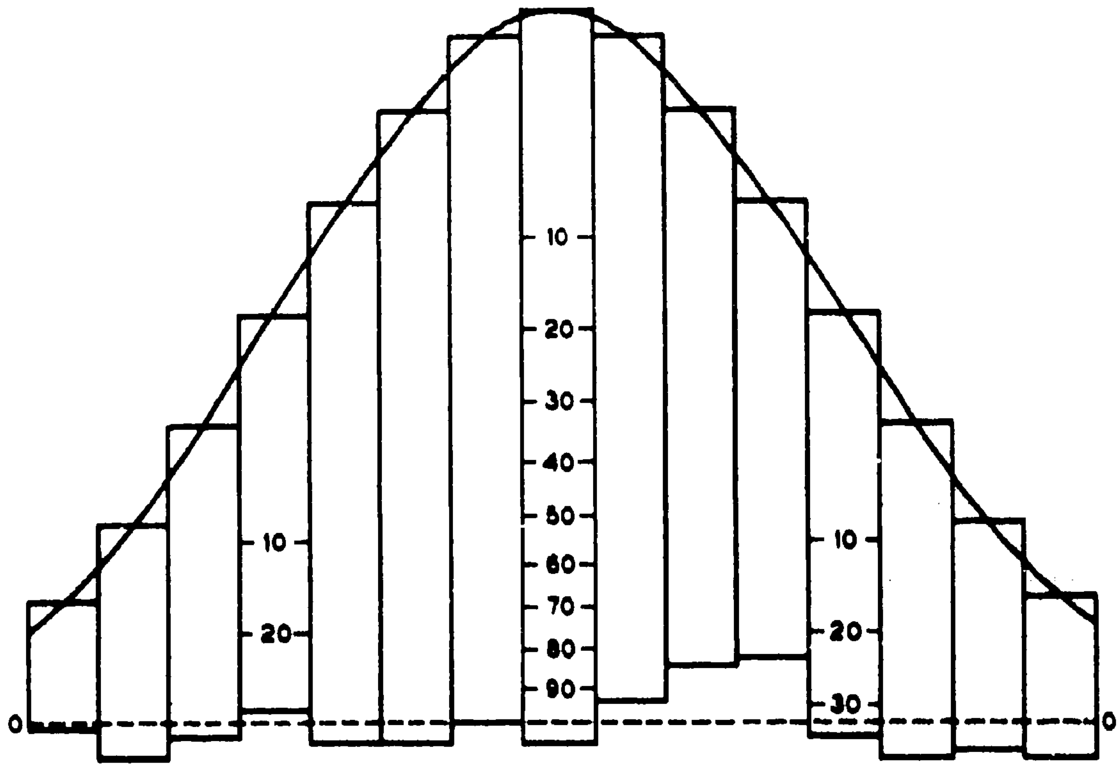


FIG. 3
HANGING ROOTOGRAM

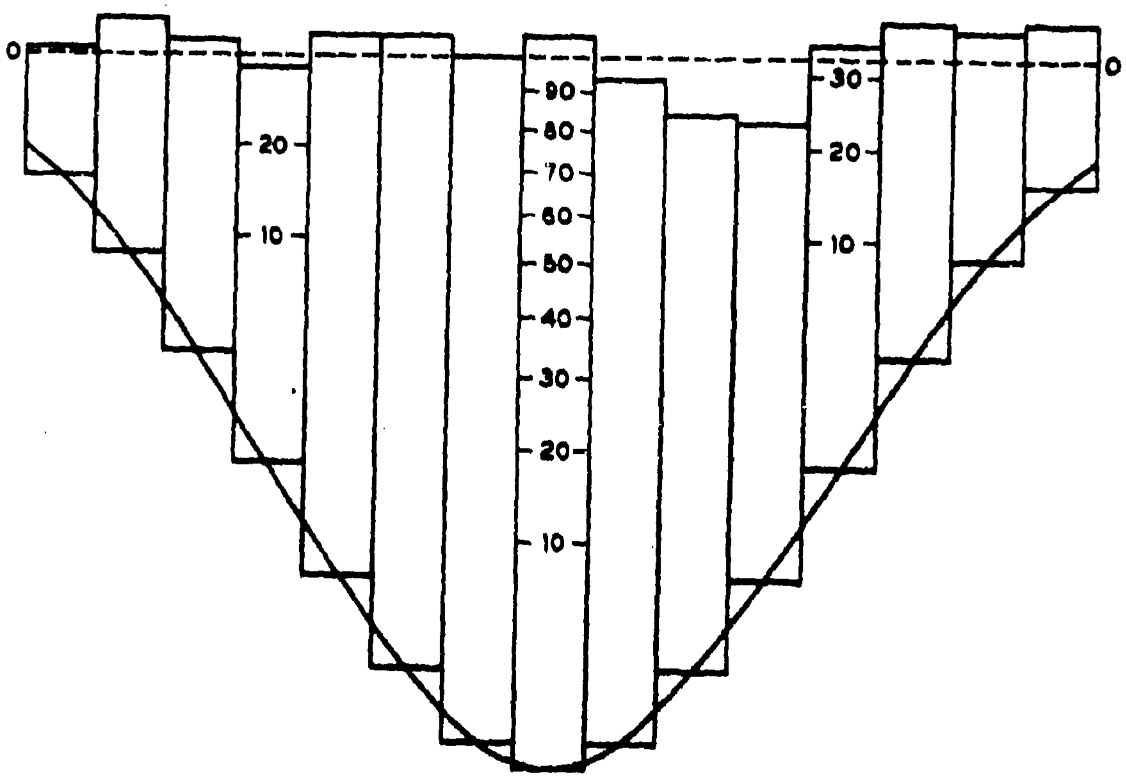


FIG. 4
SUSPENDED ROOTOGRAM

MONTE CARLO TECHNIQUES TO EVALUATE EXPERIMENTAL DESIGN ANALYSIS

M. M. Everett, D. L. Colbert, and L. W. Green, Jr.
Pratt-Whitney Aircraft
Florida Research and Development Center
West Palm Beach, Florida

1. INTRODUCTION. Monte Carlo simulation, using today's high speed computers, has opened new fields in analysis of systems previously unavailable to engineers and mathematicians. Especially valuable are the approaches they offer to study the accuracy and precision of some of the empirical relationships used in analysis of variance and experimental design.

This report gives the results of two such simulation programs made at Pratt & Whitney Aircraft's Florida Research & Development Center. It is not the purpose of this paper to present the findings of these studies as absolute truisms. They are, however, provided as the results of case histories and do offer a method for further exploration of analytical solutions in the field of analysis of variance and experimental design.

The two simulations presented here are Rejection Criteria for Approximate Student's "t" Test, and Bias in the Analysis of Variance Components from an Unbalanced Design.

2. DISCUSSION.

A. REJECTION CRITERIA FOR APPROXIMATE STUDENT'S "T" TEST

One of the simplest designed experiments is that designed to test, or compare, the first moments of two lots or populations. Of interest here is the case of the comparison of two means ($\overline{X}$ and $\overline{Y}$) calculated from the samples drawn from those populations when the population variances, σ_x^2 σ_y^2, are not equal.

This paper reports on an investigation of three different commonly used methods to determine critical values for this situation. The purpose of the investigation was to compare the relative merits of each,

assuring that the true level of significance was at least as great as the pre-selected level of significance, and to obtain an unbiased estimator having a minimum variance.

To compare, by a "t" test, two means from independent samples, where it is suspected or known that the variances are unequal, the test would be:

$$\text{(a)} \qquad t_{cal} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{s_x^2/N_x + s_y^2/N_y}} \quad .$$

When the hypothesis to be tested is that $\mu_x = \mu_y$, this reduces to

$$\text{(b)} \qquad t_{cal} = \frac{\bar{X} - \bar{Y}}{\sqrt{s_x^2/N_x + s_y^2/N_y}} \quad .$$

In the case where $\sigma_x^2 \neq \sigma_y^2$, t_{cal} does not follow the student's "t" distribution with $N_x + N_y - 2$ degrees of freedom. Therefore some critical criterion, such as a modified t-distribution, must be used to judge significance.

1. Methods Used to Determine Critical Values.

Method 1 - Cochran and Cox Approximation

$$t'_\alpha(1) = \frac{\frac{s_x^2}{N_x} t_{\alpha x} + \frac{s_y^2}{N_y} t_{\alpha y}}{\frac{s_x^2}{N_x} + \frac{s_y^2}{N_y}} \quad .$$

This method utilizes a weighted mean of the tabular t values for the two samples.

Method 2 - Dixon and Massey Approximation

t'_α (2) = tabulated value of student's "t" associated with γ degrees of freedom where

$$\gamma = \frac{\dfrac{s_x^2}{N_x} + \dfrac{s_y^2}{N_y}}{\left(\dfrac{s_x^2}{N_x} \Big/ N_x + 1\right) + \left(\dfrac{s_y^2}{N_y} \Big/ N_y + 1\right)} - 2 .$$

This approximation assumes that the mean comparisons follow a student's "t" distribution not at $(N_x + N_y - 2)$ degrees of freedom but rather at some γ degrees of freedom.

Method 3 - Satterthwaite-Welch Approximation

t'_α (3) = tabulated value of student's "t" distribution with γ degrees of freedom where

$$\gamma = \frac{\left(\dfrac{s_x^2}{N_x} + \dfrac{s_y^2}{N_y}\right)^2}{\left(\dfrac{s_x^2}{N_x}\right)^2 \Big/ (N_x - 1) + \left(\dfrac{s_y^2}{N_y}\right)^2 \Big/ (N_y - 1)} .$$

This is the approximation for the modified degrees of freedom. It is shown in various texts in different algebraic forms.

In determining t'_α (2) and t'_α (3) it is not necessary to round the degrees of freedom to the lower value; the tables can be interpolated for an unbiased estimate. However, in tables I and II, discussed later, for t'_α (2) and t'_α (3) the lower rounded degrees of freedom were used or the percentiles shown would have been somewhat smaller.

2. Simulation Procedure

These three methods were compared by Monte Carlo simulation on the IBM 7090 and 1620 computers. For a stated set of parameters and sample size, 10,000 samples were drawn from both the X and the Y populations. These samples were randomly paired and their first moments input to equation (b). The output was 10,000 values of t'_{cal}, under the restriction that $E(\overline{X} - \overline{Y}) = 0$. Then all 10,000 values were ranked and the percentiles identified. This process was repeated for various combinations of σ_x^2, σ_y^2, N_x, and N_y.

Approximate rejection values for $t'_\alpha(1)$, $t'_\alpha(2)$, and $t'_\alpha(3)$ were calculated and averaged for α at levels of 90%, 95%, and 99%. The rejection values were then compared to the appropriate percentile level from the ranked values.

3. Conclusions and Discussion.

Tables I and II summarize the test cases that were simulated. The recorded levels for the $t'_{\alpha(i)}$ are the average estimates for the actual levels 90%, 95%, and 99%. Table I is used to demonstrate the output of the simulation process. It compares the three methods when used with equal variances. The only significant conclusion demonstrated is that the Cochran & Cox approximation is an estimator whose confidence level is at least as high as the prior selected confidence level. Table II continues to demonstrate this.

If a bias exists in Method 2 and Method 3 it does so only at certain levels of the parameters and their sample sizes. This indicates that an interaction of the variables exists. Table II compares only a few situations and is entirely too general to draw many exact conclusions. It does, however, give an insight into the comparative accuracies involved.

In additional simulation studies it has been found in all cases observed that the Satterthwaite-Welch Method was a more precise estimator than the Dixon and Massey approximations. Further studies are required to obtain an unbiased estimate with a minimum variance.

It may be noted that an exact solution due to H. Scheffé has been omitted from this study. Scheffé's method is based on the fact that if $n_1 \leq n_2$, a sample of size n_1 may be randomly selected from the larger sample size of size n_2. It is then possible to calculate Scheffé's "t" statistic by:

$$t'_\alpha(4) = \frac{\left(\sum_{i=1}^{n_1} \frac{X_{1i}}{n_1} - \mu_1\right) - \left(\sum_{i=1}^{n_2} \frac{X_{2i}}{n_2} - \mu_2\right)}{\sqrt{\dfrac{\sum_{i=1}^{n_1}\left[\left\{X_{1i} - \left(\frac{n_1}{n_2}\right)^{1/2} X_{2i}\right\}^2 - \left\{\sum_{i=1}^{n_1}\left[X_{1i} - \left(\frac{n_1}{n_2}\right)^{1/2} X_{2i}\right]\right\}^2 \Big/ n_1\right]}{n_1\,(n_1-1)}}}$$

which is distributed as Student's "t" with n_1-1 degrees of freedom. It is immediately obvious that the relative information of this statistic decreases as the value of $n_2 - n_1$ becomes large, since $n_2 - n_1$ observations are randomly eliminated from the calculation of the "t" statistic under the assumption that $n_1 \leq n_2$.

Since this loss of information is especially severe for the case where one sample size is very small, this method was not considered in this study.

B. BIAS IN THE ANALYSIS OF VARIANCE COMPONENTS FROM AN UNBALANCED DESIGN.

It was suspected that a bias existed in the estimates of the components of variance when analysis of variance techniques are applied. It was further suspected that this bias was due to unequal sample sizes. If this bias could be related to sample sizes and sample size ratios then it may be possible to derive an unbiasing technique. Using the IBM 7090 computer, a Monte Carlo Simulator was written to determine if this suspected bias existed and to study the possibilities of finding a method to identify this bias.

1. Simulation Procedure

The simulator was designed to determine the distribution of estimates of process variance from a one-way ANOVA, byproduct of which were estimates of the within-process variation discussed below. To simulate the two sources of variation, two populations of normally distributed random numbers were set up. The means of these distributions were given fixed arbitrary values, while the standard deviations (and therefore, the variances) were variable. The first population was designated as the process-to-process source of variation. The second was designated as the within-process source of variation. It was decided that three ratios of standard deviations of these populations would be used. There were:

$$\frac{\sigma \text{ process-to-process}}{\sigma \text{ within-process}} \text{ of } 0.5,\ 1.0,\ 2.0 \quad .$$

These values were selected because they cover the general area of interest in estimating process-to-process variation. It was further decided that for each ratio above, a control case (balanced data) should be run, in addition to a case with mild unbalance, and a case with extreme unbalance. The balanced case had four runs with 10 data points for each run. The mildly unbalanced case had four runs with 8, 9, 10, and 12 data points each. The extremely unbalanced case had four runs with 5, 3, 10, and 15 data points in each. The analysis of variance described above was carried out for each case on an IBM 7090 computer in the following sequence:

1. Four values were selected from the population of process-to-process random numbers. One of these corresponds to each process.

2. Four sets of numbers were selected from the within-process population. Each of these sets corresponded to one of the four processes. For example, for the control case, each set would have ten random numbers; for the extremely unbalanced case, the set corresponding to the first process would have five members, the second set would have three, etc. The value, selected in step 1, for the first process is added to each member of the set of within-process numbers for the first process; the

second process number is added to each value of the second set, etc. In this manner an array of numbers is produced. Each column has a common process effect while within each column there is a within-process effect.

3. The formulas of the analysis of variance were used to estimate the process-to-process and within-process variances. These values are stored in the computer.

4. Steps 1-3 are repeated 1000 times so that the 1000 estimates of each variance are obtained.

5. The 1000 values are then ranked and the mean, standard deviation, and standard error of the mean are computed. The ranked estimates and computed statistics are then listed.

6. The plot, figure 1, was then made, showing the frequency distribution of the estimates.

For comparison purposes the plots of the control case, the mild unbalanced case, and the extremely unbalanced case are shown together in figure 1 for the ratio 1.0 to 1.0 of standard deviations.

The results of the simulation are summarized in table III. If a bias exists in either the process-to-process or the within-process variance it is not evident here. There is however, a relatively large scatter of the variance estimates.

2. Conclusions and Discussion

Based on the Monte Carlo Simulator, the following conclusions were reached:

1. The distributions of the estimates of within-process variances (discussed in the appendix) were approximately normal. These distributions exhibited a marked central tendency and a degree of symmetry.

2. The bias in the estimates of within-process variability was negligible in each case tested. The cases using unbalanced

data (unequal sample sizes within run) did not demonstrate biases significantly larger than the control case.

These first conclusions were not unexpected and were more or less byproducts of the simulation. The more important conclusions follow:

3. The distributions (figure 1) of the estimates of process-to-process variances were highly skewed to the right and truncated on the left. However, the cases using the unbalanced data were no more skewed than the control case (balanced data).

4. The bias in the estimates of process-to-process variability was negligible in each case tested, including the cases of unbalanced data.

5. Although estimates of process-to-process variance resulting from the analysis of variance technique are not optimum, no known method of improving this situation exists.

The last three conclusions presented represent the main intent of this study. It must be noted that the estimates of process-to-process variance cannot be considered "optimum" estimates. Since an optimum estimate should have minimum variance and minimum total error, the estimates based on analysis of variance techniques cannot be optimized. Any attempt to further optimize them through the use of an unbiasing technique must reduce the variance of the estimates and, at the same-time, increase the relative value of each estimate (because of skewness).

In figure 2 the sample curve for the distribution of estimates of process-to-process variance is skewed and exhibits a large amount of scatter. The distribution of an optimum estimating procedure should have minimum variance and minimum bias, approached by the second curve shown in figure 2. To optimize the present estimates (obtained from ANOVA) of process-to-process variance, the scatter of the distribution of these estimates should be reduced. To accomplish this, each estimate (s^2) should be divided by a factor K, where K is greater than one (1). The proper selection of K will minimize the variance. However, this will bias the estimate so that a correction must be made; that is, the mean estimate of the variance will be only 1/K of the true value. Thus, 1 - 1/K must be added to each estimate s^2 to unbias the estimates. However for any estimate s^2;

$$(1/K)s^2 + (1-1/K)s^2 = s^2/K + s^2 - s^2/K = s^2 .$$

Therefore, for any K selected the estimate is not improved. Any other plan to optimize these estimates will fail since to reduce the scatter a bias must be introduced and to minimize this bias the scatter must be increased.

REFERENCES

1. Anderson, R. L., and Bancroft, T. A. Statistical Theory in Research. New York: McGraw-Hill Book Company, Inc., 1952

2. Bowker, A. H. and Lieberman, G. J. Engineering Statistics. New Jersey: Prentice-Hall, Inc., 1959

3. Brownlee, K. A. Statistical Theory and Methodology in Science and Engineering. New York: John Wiley and Sons, Inc., 1960

4. Cochran, W. G. and Cox, G. M. Experimental Design. New York: John Wiley and Sons, Inc., 1957

5. Davies, Owen L. Statistical Methods in Research and Production. New York: Hafner Publishing Company, 1961

6. Dixon, W. J. and Massey, F. J., Jr. Introduction to Statistical Analysis New York: McGraw-Hill Book Company, 1957.

7. Duncan, A. J. Quality Control and Industrial Statistics. Homewood, Illinois: Richard D. Irwin, Inc., 1959

8. Hald, A. Statistical Theory with Engineering Applications. New York: John Wiley and Sons, Inc., 1962

9. Steel, R. G. D. and Torrie, J. H. Principles and Procedures of Statistics New York: McGraw-Hill Book Company, Inc., 1960

TABLE I

EQUAL VARIANCES

	N_X	N_Y	σX^2	σY^2	Simulated Actual α Level	t'_α (1)	t'_α (2)	t'_α (3)
Case I	10	10	1	1	90	90.8	$\underline{89.1}$	$\underline{89.1}$
	10	10	1	1	95	95.8	$\underline{94.0}$	$\underline{94.2}$
	10	10	1	1	99	99.5	$\underline{98.7}$	$\underline{99}$
Case II	5	3	1	1	90	94.6	$\underline{89.1}$	92.6
	5	3	1	1	95	98	$\underline{95.4}$	95
	5	3	1	1	99	99.8	$\underline{97.8}$	$\underline{98.2}$
Case III	15	8	1	1	90	90.7	90.6	90.4
	15	8	1	1	95	95.6	$\underline{94.1}$	95.2
	15	8	1	1	99	99.2	$\underline{99.1}$	99.1

All α levels expressed in percent;
Underlined values denote underestimates

TABLE II

UNEQUAL VARIANCES

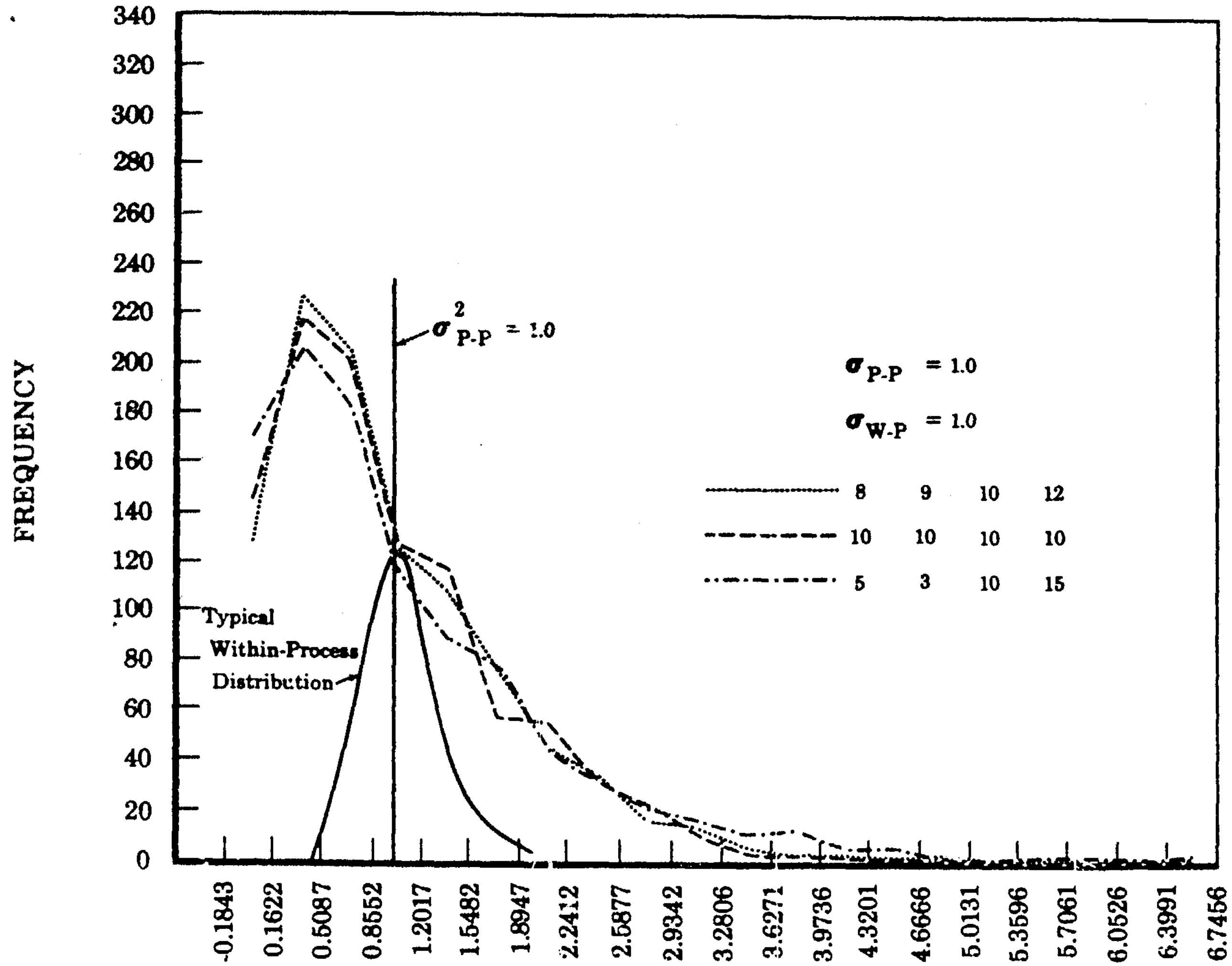
	N_X	N_Y	σX^2	σY^2	Simulated Actual α Level	t'_α (1)	t'_α (2)	t'_α (3)
Case I	15	5	9	1	90.	93.6	92.1	92.0
	15	5	9	1	95.	96.4	95.3	95.3
	15	5	9	1	99.	99.3	97.9	98.6
Case II	5	15	16	1	90.	94	95	95
	5	15	16	1	95.	97	97.3	97.3
	5	15	16	1	99.	99.8	99.4	99.4
Casc III	15	8	.0625	1	90.	92	92	92
	15	8	.0625	1	95.	96.7	96.7	96.7
	15	8	.0625	1	99.	99.8	99.8	99.8
Case IV	12	8	2.25	1	90	91.8	88.1	89.4
	12	8	2.25	1	95	96.0	94.3	94.4
	12	8	2.25	1	99	99.3	98.6	98.6
Case V	5	3	4	1	90	91.4	86.1	86.9
	5	3	4	1	95	95.3	92.0	92.7
	5	3	4	1	99	99.6	97.6	98.1
Case VI	30	18	2.25	1	90	97	93	93
	30	18	2.25	1	95	99.7	97	97
	30	18	2.25	1	99	99.9	98.9	98.9

All α levels expressed in percent;
Underlined values denote underestimates

TABLE III

	Inputs			Estimates		Variance of the Variance Estimate About its Mean	
No. of Process	Process-Process Variance	Within-Process Variance	Numbers of Observations within each Process	Process-Process Variance	Within-Process Variance	Process-Process	Within-Process
4	1.0	0.25	8, 9, 10, 12	0.9876	0.2524	0.7381	0.0037
4	1.0	0.25	10, 10, 10, 10	1.0219	0.2543	0.6824	0.0037
4	1.0	0.25	5, 3, 10, 15	0.9588	0.2496	0.8566	0.0044
4	1.0	1.0	8, 9, 10, 12	0.9812	1.0069	0.7564	0.0558
4	1.0	1.0	10, 10, 10, 10	0.9418	0.9999	0.7086	0.0538
4	1.0	1.0	5, 3, 10, 15	1.0297	1.0163	1.0658	0.0657
4	0.25	1.0	8, 9, 10, 12	0.2569	1.0037	0.0854	0.0609
4	0.25	1.0	10, 10, 10, 10	0.2353	1.0057	0.0732	0.0568
4	0.25	1.0	5, 3, 10, 15	0.2624	0.9903	0.1127	0.0657

Process-to-Process Components of Variance
1000 Observations

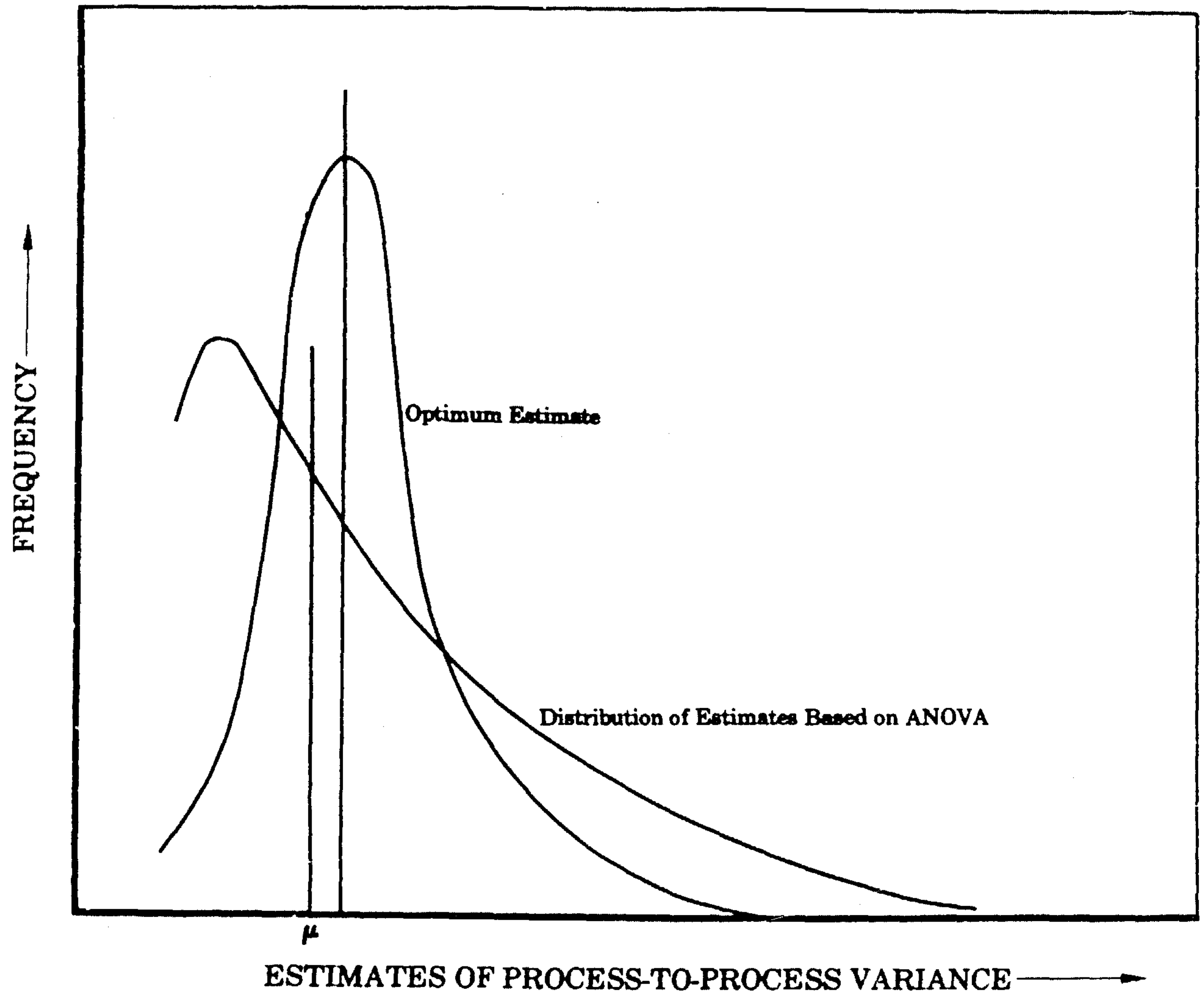


DIVISION OF UNITED AIRCRAFT CORPORATION

FIGURE I

643010
FD9799

Frequency Distribution, Process-to-Process Variance



643010
FD9800

DIVISION OF UNITED AIRCRAFT CORPORATION

FIGURE II

LIST OF ATTENDEES

Addelman, Dr. Sidney	Research Triangle Institute
Ammann, Mr. W. H.	US Army Aviation Materiel Command
Atkinson, Mr. John C.	CRDL, Edgewood Arsenal
Ayres, Mr. James N.	US Naval Ordnance Lab
Bartee, Mr. Edwin M.	Univ of Alabama
Bartko, Dr. John J.	Natl Inst of Mental Health
Bartlett, Mr. Richard P., Jr.	US Dept of Agriculture
Beall, Mr. John R.	US Army Medical R&D Command
Bechhofer, Prof Robert	Cornell Univ
Bercaw, Maj W. W.	US Army Strategy & Tactics Analysis Gp.
Berndt, Mr. Gerald D.	Hq SAC, Offutt AFB
Betts, Maj Genl. Austin W.	Deputy Chief of Res. & Devel., D/A
Box, Prof George E. P.	Univ of Wisconsin
Boyle, Mr. Douglas G.	Dugway Proving Ground
Bright, Mr. Jerry W.	CRDL, Edgewood Arsenal
Brinkmann, Mr. George L.	FDA, Washington, D. C.
Brown, Mr. Ralph E.	Frankford Arsenal
Brown, Mr. Wm. A.	Dugway Proving Ground
Bruno, Mr. O. P.	Army Ballistic Res Lab
Bryson, Dr. Marion R.	Duke Univ
Bulfinch, Mr. Alonzo	Picatinny Arsenal
Burdick, Dr. Donald S.	Duke Univ
Burke, Col James L.	Ft. Huachuca
Bustead, Ronald	US Army Natick Lab
Byas, Mr. W. E.	Picatinny Arsenal
Cameron, Dr. Jos. M.	Stat Engrg Lab, NBS
Carbonaro, Philip A. G.	US Army Materials Res. Agency
Carter, Mr. Frederick L., Jr.	Ft. Detrick
Christianson, Mr. C. J.	RAC
Ciuchta, Mr. H. P.	CRDL, Edgewood Arsenal
Coffman, Rebecca J.	Eng. R&D Lab.
Cogdell, LCdr. John J.	COMOPTEVFOR
Cohen, Prof. A. C., Jr.	Univ of Georgia
Coleman, Mr. Roger D.	Johns Hopkins Univ
Cook, Mr. Charles M., Jr.	Opera. Eval. Group
Coon, Helen J.	Ballistic Res Lab
Cox, Mrs. Claire B.	NIH
Cox, Dr. Edwin L.	USDA
Crancer, Mr. Alfred, Jr.	AF Office of Scientific Res.

Cox, Mr. Paul C.	White Sands Missile Range
Craw, Dr. Alexander R.	Fort Detrick
Curtis, Mr. John J.	Fort Detrick
Cutchis, Mrs. Angeliki D.	Johns Hopkins Univ
Danish, Mr. Michael B.	BRL, Aberdeen
David, Prof H. A.	Univ of N. C.
DeArmon, Mr. Ira A., Jr.	ORG, Edgewood Arsenal
DeCicco, Mr. Henry	USA MUCOM
Demchak, Mr. Peter	Fort Detrick
Dihm, Mr. Henry, Jr.	Redstone Arsenal
Dimling, Lt. John A.	US Army Strategy & Tactics Analysis Gp.
Dobrindt, Mr. Gerald T.	Analytical Lab, Aberdeen
Dressel, Dr. F. G.	ARO-D, Durham
Duncan, Dr. David B.	Johns Hopkins Univ
Dunn, Mr. Paul F.	Booz-Allen
Eisenhart, Dr. Churchill	NBS
Eissner, Mr. Robert M.	BRL, Aberdeen
Emero, Mr. Roland F.	Raytheon Company
Ellerson, Mrs. Elizabeth C.	US Naval Propellant Plant
Ellner, Mr. Henry	Directorate for Quality Assurance, EA
Endelman, Miss Anna	Food & Drug Admin, Bur of Reg. Compl.
Engel, Mr. Klaus H. C.	AERO Space Div
Enis, Mr. Peter	George Washington Univ
Ewart, Mr. Wade H.	US Army Missile Command
Fichter, Mr. Lewis S.	Picatinny Arsenal
Fiddleman, Capt. Paul B.	USCRDL, Edgewood
Foster, Dr. Walter D.	Fort Detrick
Freedle, Dr. Roy O.	Amer. Inst. for Res.
Frishman, Mr. Fred	ARO, OCRD, DA
Galvin, Mr. Cyril J., Jr.	Coastal Engrg Res Center
Galbraith, Dr. A. S.	ARO-D, Durham
Gardner, Dr. Roberta A.	Div of Biol. Stds., NIH
Gehan, Dr. Ed	Natl. Cancer Inst.
Geisser, Dr. Seymour	NIAMD
Glick, Mr. Charles E.	Atlantic Res Corp.
Gliser, Leon J.	Columbia Univ
Goldstein, Col. J. D.	USA R&D Command
Graesel, David B.	Natl Cash Register Co.
Granville, Mr. William, Jr.	Frankford Arsenal

Green, Mr. Lee, Jr.	Pratt-Whitney Aircraft
Green, Mr. Lyle D.	Inst of Sc. & Tech, U of Mich.
Green, Mr. Theron D.	Ft. Detrick
Greenhouse, Dr. Samuel W.	Natl. Inst. of Mental Health
Greenwood, Dr. Joseph A.	Naval Medical Center
Grimes, Mr. James Paskell	Naval Res Lab
Grubbs, Dr. Frank E.	Ballistic Res Labs, Aberdeen
Haberman, Mr. Sol	Johns Hopkins Univ
Haines, Dr. Bertram W.	Ft. Detrick
Hampton, Mr. L. D.	U. S. Naval Ordnance Lab.
Hanson, Dr. Fred S.	White Sands Missile Range
Harris, Dr. Bernard	Math. Res. Center, Madison, Wis
Hartley, Prof. H. O.	Inst. of Stat., Texas A&M Univ
Hassel, Mr. L. D.	Picatinny Arsenal
Hauser, Mr. Arthur L.	FDA, Washington, D. C.
Heaney, Mrs. Marian	Dept of Defense, Ft. Meade
Hecht, Mr. Edward C.	Picatinny Arsenal
Helvig, Mr. T. N.	Honeywell Inc.
Hershner, Dr. Ivan R.	ARO, OCRD, DA
Hess, Mr. Martin	Koppers Company
Hixson, Mr. Eugene E.	GSFC-NASA, Greenbelt
Holmes, Lt Col. Frederick B.	USAR, R&D Unit, Sacramento, Cal.
Homeyer, Mr. Paul G.	CEIR
Hook, Mr. Jonn R.	Hq, DASA
Hoppes, Mr. Harrison N.	RAC, McLean, Va.
Horner, Dr. Theodore W.	Booz-Allen
Howes, Mr. David R.	US Army Strategy & Tactics Analysis Gp
Inselmann, Dr. Edmund H.	Frankford Arsenal
Isaac, Mr. Gerhard J.	US Army Med Res & Nutritions Lab
Itkin, Mr. Arthur	Merck Sharp & Dohme Res Lab
Jacobus, Dr. David P.	WRAIR
Jebe, Prof. Emil H.	Univ of Michigan
Jenkins, Mr. Andrew H.	Redstone Arsenal
Jessup, Dr. Gordon L.	Ft. Detrick
John, Mr. Frank J.	Watervliet Arsenal
Johnson, Mr. Cecil D.	USAPRO
Johnson, Mr. Jerome R.	BRL, Aberdeen
Johnson, Mr. Robert S.	FDA, Wash. D. C.
Jones, Mrs. Marian W.	Ft. Detrick
Jones, Mr. Richard	Johns Hopkins Univ

Karp, Mr. A. E.	US Army Strategy & Tactics Analysis Gp
Katz, Dr. Darryl	Douglas Aircraft
Kaufman, Mr. J. V. Richard	USA MUCOM
Kempthorne, Prof Oscar	Iowa State Univ
Kendall, Dr. M. G.	CEIR
Kiefer, Prof Jack C.	Cornell Univ
Killion, Dr. Lawrence E.	Ft. Huachuca, Ariz
Kinsinger, Mrs. Pauline B.	TRECOM, Ft. Eustis, Va.
Kirby, Dr. Wm. H., Jr.	BRL, Aberdeen
Kirby, Mr. Michael	RAC, McLean, Va.
Knetz, Dr. Wallace J.	Amer. Inst for Res.
Kniss, Mr. James R.	Ballistic Res Labs, Aberdeen
Kokinakia, Mr. Wm.	TBL, BRL, Aberdeen
Kramer, Dr. Clyde Y.	VPI
Krimmer, Mr. Manfred W.	Ammunition Proc & Supply Agency
Kroll, Mr. Wm. F.	Johns Hopkins Univ
Kruse, Mr. Richard H.	Ft. Detrick
Ku, Mr. Hsien H.	Stat Engrg Lab, Natl Bur of Stds
Kupperman, Dr. Morton	Natl. Security Agency, Ft. Meade
Kurkjian, Dr. Badrig	Harry Diamond Labs
Kutger, Dr. Gerald	US Army Strategy & Tactics Analysis Gp
Lane, Mr. Joseph R.	US Army Res Office, Durham
Lawler, Mr. John M.	US Army Natick Labs
LeClerg, Dr. E. L.	Biometrical Svcs, ARS
Lee, Dr. C. Bruce	Science & Tech Div, Library of Congress
Lehman, Dr. Alfred	WRAIR
Lerche, Kenneth D.	Ft. Meade
Lieberman, Prof Gerald J.	Stanford Univ
Lieblein, Mr. Julius	US Post Office Dept
Lieberman, Mr. Herbert J.	Bur of Supplies & Accts
Little, Mr. Robert E.	Oklahoma Stat Univ
Loe, Miss H. V.	Bu of Ships, Navy Dept
Long, Mr. Melvin H.	Frankford Arsenal
Lowery, Mr. Earl D.	US Army Strategy & Tactics Analysis Gp.
Lucas, Prof H. L.	NC State College
Lundegard, Dr. Robert	Office of Naval Res
Lundy, Hazel L.	Springfield Armory
Madden, Mr. Dale A.	Atlantic Res. Corp.
Mall, Mr. Adolph W.	US Army Strategy & Tactics Analysis Gp.

Malligo, Mr. John E.	Ft. Detrick
Maloney, Dr. Clifford J.	NIH
Mandelson, Mr. Joseph	Dir. of Quality Assurance, Edgewood
Mann, Dr. Henry B.	Math Res Center, USA, Univ of Wis
Manthei, Mr. James H.	CRDL, MED RES, Edgewood
Martin, Mr. Francis F.	Booz Allen
Masaitis, Dr. C.	BRL-Aberdeen
Matthews, Mr. Gerald A.	Mississippi State Univ
Matthis, Mr. Carlton L.	Cornell Aeronautical Lab
Mauss, Mrs. Besse Day	Wash., D. C.
McCormick, Mr. Garth	Research Analysis Corp
McGroddy, Miss Patricia	STAG
McIntosh, Mr. Albert L.	Ft. Huachuca, Ariz
McIntosh, Dr. Wm. B.	Ft. Huachuca, Ariz
Menken, Mrs. Jane A.	Natl Inst of Mental Health
Miller, Mrs. Christine	NIH
Mood, Dr. A. M.	US Office of Education
Moshman, Mr. Jack	CEIR Corp
Moss, Mr. David M.	Booz-Allen
Moss, Mr. H. Donald	West. Elec. Corp
Myers, Mr. Raymond H.	VPI
Mylander, Mr. W. Charles	Research Analysis Corp
Nassimbene, Mr. Raymond	US Bureau of Budget
Natrella, Mrs. Mary G.	Stat. Engrg Lab, NBS
O'Connor, Mr. Desmond	GIMRADA, Ft. Belvoir
Oklin, Prof Ingram	Stanford Univ
Olon, Mr. Frederick A.	Naval Propellant Plant
Orleans, Beatrice S.	BuShips
Pabst, Dr. Wm. R., Jr.	Bureau of Naval Weapons
Panos, Mr. Robert	CEIR
Parrish, Dr. Gene B.	ARO-Durham
Parvin, Mr. David W., Jr.	Mississippi State Univ
Parzen, Dr. Emanuel	Stanford Univ
Pepper, Mr. Leonard	USAE Waterways Experiment Station
Perry, Virginia W.	Ft. Lee, Va.
Persweig, Mr. Michael	Picatinny Arsenal
Pettigrew, Mr. Hugh	Natl. Cancer Institute
Pfeiffer, Mr. Otto H.	US Army Tank-Automotive Center
Piepoli, Mr. Carl R.	Ft. Detrick
Podolsky, Mr. Benjamin	Frankford Arsenal

Pollock, Mr. Abraham	OCRD, DA
Quarles, Dr. Gilford G.	Office, Chief of Engrg
Ravenis, Mr. Jos. V. J., II	Johns Hopkins Univ
Rhian, Mr. Morris A.	ORG, Edgewood
Richardson, Mr. B. A.	Canadian Army Opera Res Establishment, Ottawa, Canada
Rider, Dr. Paul R.	Aerospace Res Lab, Wright-Patterson AFB
Riggs, Mr. Charles W.	Ft. Detrick
Riley, Mr. Donald C.	Amer Stat. Assoc
Roberts, Dr. Charles	NIH
Roberts, Mr. Sean C.	Mississippi State Univ
Roetzel, Mr. Thomas G.	Ft. Detrick
Rhode, Dr. Charles A.	Johns Hopkins Univ
Rosenblatt, Dr. H. M.	Bureau of the Census
Rosenblatt, Dr. Joan R.	Natl Bureau of Standards
Ross, Mr. Alan	Johns Hopkins Univ
Roth, Dr. Raymond E.	St. Bonaventure Univ
Rotkin, Mr. Israel	Harry Diamond Labs
Rust, Mr. Philip G.	Winnstead Plantation
Sathe, Dr. Y.	Natl. Cancer Inst.
Schilling, Col. Charles H.	US Military Academy, West Point, N. Y.
Schenker, George	Army Weapons Command
Schmidt, Dr. Th. W.	US Army Res, Durham
Schalten, Roger W.	Boeing Co.
Selig, Mr. Seymour M.	Office of Naval Research
Selman, Mr. Jerome H. N.	MUCOM
Simon, Maj Gen Leslie E. (Ret'd)	Winter Park, Fla.
Sirota, Mr. Milton	Defense Supply Agency
Smith, Dr. Eugene F.	Waterways Experiment Sta.
Snyder, Mr. Mitchell	Univ of Chicago
Soland, Dr. Richard	Research Analysis Corp.
Soni, Mr. Atmaram H.	Oklahoma State Univ
Sorenson, Mr. Richard C.	US Army Personnel Res Office
Speckman, Miss Janace A.	Natl Bureau of Standards
Starr, Dr. Selig	Army Res Office, OCRD, DA
Stearman, Dr. Robert L.	CEIR
Steedman, Mr. Joseph E.	USA CDCOA Aberdeen
Stephanides, Mrs. Agath S.	US Army Strategy & Tactics Analysis Gp.
Sutherland, William H.	Res. Analysis Corp.
Tallis, Mr.	Johns Hopkins Univ

Tank, Capt. Douglas B.	Walter Reed Army Inst. of Res.
Tingey, Lt. Henry B.	Ballistic Res Labs
Trethewey, John D.	DTC
Tukey, Prof. John W.	Princeton Univ
Vick, Mr. James A.	CRDL, Edgewood Arsenal
VonGuerard, Dr. Herman W.	FMC Corp
Wadley, Dr. F. M.	Ft. Detrick
Walner, Arthur H.	US Navy Applied Science Lab
Wallach, Mr. Harold	Public Housing Adm.
Wampler, Roy H.	Natl. Bureau of Standards
Watson, Prof. G. S.	Johns Hopkins Univ
Watson, Mr. H. H.	Dept of Natl Defense, Ontario, Canada
Weingaiten, Dr. Harry	Dept of Commerce
Weinstein, Mr. Joseph	Ft. Monmouth
Weintraub, Gertrude	Picatinny Arsenal
Wenger, Mr. Warren	Ballistic Res Labs, Aberdeen
Westrope, Mr. John	US Army Strategy & Tactics Analysis Gp.
Weyland, Maj Carl E.	Hq, OAR
Wicham, Mr. Robert E.	Picatinny Arsenal
Wigler, Kenneth	Naval Command Systems Support Activity
Wisenfeld, Mr. L.	Picatinny Arsenal
Williams, Henry K.	CRDL, Edgewood
Williams, Mr. Jacob A.	US Army Strategy and Tactics Analysis Gp.
Williams, John N.	United Aircraft Corp.
Willke, Dr. Thomas A.	Natl Bur of Standards
Wolman, Dr. Wm.	NASA-Goddard Space Center
Woodal, Mrs. Rosalie C.	HDL
Youden, Dr. W. J.	Natl Bureau of Standards
Young, Mr. Harold W.	Ft. Detrick
Zelen, Dr. Marvin	Natl. Cancer Inst.
Zimmerman, Mr. J. M.	Rocketdyne
Zundel, Brown	CEIR
Zweifel, Mr. Jim	Natl. Cancer Inst.